Coarse-Graining Based on Pair Interactions – Studies on Transferability and Dynamic Consistency in Coarse-Grained Models of Soft Matter

vom Fachbereich Chemie
der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium
(Dr. rer. nat.)

Dissertation
von Gregor Deichmann

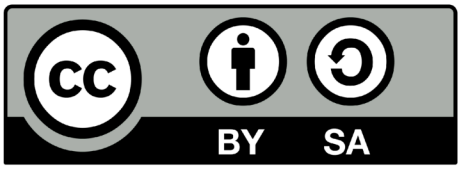
Erstgutachter: Prof. Dr. Nico F. A. van der Vegt
Zweitgutachter: Prof. Dr. Florian Müller-Plathe

Darmstadt, 2019

Tag der Einreichung: 02. September 2019

Tag der mündlichen Prüfung: 14. Oktober 2019

URN: urn:nbn:de:tuda-tuprints-92607



Erklärung zu Selbstständigkeit und Promotionsversuch

Darmstadt, den

Ich erkläre hiermit, dass ich diese Dissertation selbstständig nur unter Zuhilfenahme der angegebenen Hilfsmittel verfasst habe und dass ich noch keinen Promotionsversuch unternommen habe.

(Deichmann, Gregor)

Erkärung der Übereinstimmung

Darmstadt, den

Ich erkläre hiermit, dass die elektronische Version dieser Dissertation mit der schriftlichen Fassung übereinstimmt. Die elektronische Version liegt dem Prüfungssekretariat vor.

(Deichmann, Gregor)

Erklärung zum Eigenanteil der Veröffentlichungen

Teile dieser Arbeit wurde in folgenden Artikeln veröffentlicht. Ich war an den Veröffentlichungen mit folgendem prozentualem Anteil beteiligt:

1. **G. Deichmann**; M. Dallavalle; D. Rosenberger; N. F. A. van der Vegt, Phase Equilibria Modeling with Systematically Coarse-Grained Models - A Comparative Study on State Point Transferability, *Journal of Physical Chemistry B*, **2019**, 123 (2), p. 504-515
   Eigenanteil 80%

2. **G. Deichmann**; N. F. A. van der Vegt, Conditional Reversible Work Coarse-Grained Models of Molecular Liquids with Coulomb Electrostatics - a Proof of Concept Study on Weakly Polar Organic Molecules, *Journal of Chemical Theory and Computation*, **2017**, 13 (12), p. 6158-6166
   Eigenanteil 100%

3. **G. Deichmann**; N. F. A. van der Vegt, Conditional Reversible Work Coarse-Grained Models with Explicit Electrostatics - An Application to Butylmethylimidazolium Ionic Liquids, *Journal of Chemical Theory and Computation*, **2019**, 15 (2), p. 1187-1198
   Eigenanteil 100%

4. **G. Deichmann**; N. F. A. van der Vegt, Bottom-up approach to represent dynamic properties in coarse-grained molecular simulations, *Journal of Chemical Physics*, **2018**, 149, p. 244114
   Eigenanteil 100%

Datum:

Unterschrift

Erklärung zur Begutachtung der Veröffentlichungen

Referent: Prof. Dr. Nico van der Vegt
Korreferent: Prof. Dr. Florian Müller-Plathe

Weder Referent (Prof. Dr. Nico van der Vegt) noch Korreferent (Prof. Dr. Florian Müller-Plathe) waren an der Begutachtung der nachstehenden Veröffentlichungen beteiligt:

1. **G. Deichmann**; M. Dallavalle; D. Rosenberger; N. F. A. van der Vegt, Phase Equilibria Modeling with Systematically Coarse-Grained Models - A Comparative Study on State Point Transferability, *Journal of Physical Chemistry B*, **2019**, 123 (2), p. 504-515
2. **G. Deichmann**; N. F. A. van der Vegt, Conditional Reversible Work Coarse-Grained Models of Molecular Liquids with Coulomb Electrostatics - a Proof of Concept Study on Weakly Polar Organic Molecules, *Journal of Chemical Theory and Computation*, **2017**, 13 (12), p. 6158-6166
3. **G. Deichmann**; N. F. A. van der Vegt, Conditional Reversible Work Coarse-Grained Models with Explicit Electrostatics - An Application to Butylmethylimidazolium Ionic Liquids, *Journal of Chemical Theory and Computation*, **2019**, 15 (2), p. 1187-1198
4. **G. Deichmann**; N. F. A. van der Vegt, Bottom-up approach to represent dynamic properties in coarse-grained molecular simulations, *Journal of Chemical Physics*, **2018**, 149, p. 244114

Datum: 5.7.19

Referent (Prof. Dr. Nico van der Vegt)

Korreferent (Prof. Dr. Florian Müller-Plathe)

Curriculum Vitae

Gregor Benedikt Deichmann, geboren am 14.11.1988 in Heidelberg.

2007	Abitur am Starkenburg-Gymnasium Heppenheim
2008-2011	Studium BSc Chemie TU Darmstadt Abschlussarbeit: "" (AK Van der Vegt)
2011-2012	Auslandsstudium an der Université de Bordeaux I
2011-2014	Studium MSc Chemie TU Darmstadt Abschlussarbeit: "" (AK Van der Vegt)
2014-2018	Promotionsstudium im Arbeitskreis Van der Vegt

Summary

Computer simulations of molecules and atoms are useful tools in soft matter research. Physical chemistry has profited significantly from the insight provided by the use of classical molecular dynamics simulations. In these simulations, atoms are modeled as point masses and their propagation in time and space is described by iterative solving of Newton's equations of motion. A major challenge lies in the fact that the computational resources required to simulate systems at this resolution render atomistical simulations prohibitively expensive already at comparatively small time and length scales. One approach to make simulations of given systems more efficient in computational terms is to coarse-grain the model, i.e., to reduce the spatial resolution by merging atoms into larger interaction sites. The procedure of coarse-graining consists of two steps: the definition of a mapping between the scales of resolution and the determination of suitable potentials for the interactions between the sites of the coarse-grained model. Especially the second step is challenging because coarse-grained model needs to reproduce the physical behavior of the underlying (fine-grained) reference model as close as possible to retain its predictive quality. Several methods which share an approach described as systematic, bottom-up coarse-graining, have been published in the literature to determine interactions in the coarse-grained model from interactions in fine-grained (mostly atomistic) reference models of the systems of interest. The transferability of a coarse-grained model, i.e, the capability of accurately reproducing predictions of the reference model at varying state points is especially dependent on the method chosen for the parameterization of the coarse-grained model. Among the existing systematic coarse-graining methods, the conditional reversible work (CRW) method achieves a high degree of transferability, while being conceptually simple, straightforward to implement and computationally efficient.

In this thesis, studies are presented which aim at an extension of the applications and systems of CRW-parameterized models. In a comparative study, the CRW method is used, among others, for the study of vapor-liquid equilibria and the thermodynamics of mixing with coarse-grained models of hexane and perfluorohexane. Results confirm the strong dependence of model transferability on the coarse-graining method chosen for its parameterization and show that the CRW models are transferable with respect to temperature, transfer from the interface to the bulk, transfer from the vapor to the liquid phase, and composition of a binary mixture. In the existing literature, the CRW method has only been applied to systems of apolar hydrocarbons. This thesis presents studies in which CRW models are parameterized for systems of weakly polar organic molecules and ionic liquids. The resulting CRW models are transferable to the same degree as those of apolar systems.

Another major challenge in the simulation with coarse-grained models is the reproduction of dynamic properties in accordance with fine-grained reference models. In general, the time scales of relaxation are smaller in coarser models and this leads to an effective 'speed-up' of these simulations, a behavior which is related to a loss of dissipative degrees of freedom in the coarser model. The effective impact of these degrees of freedom on the dynamics of the system can be simulated through the insertion of dissipative pair interactions into the coarse-grained model. These interactions can be parameterized, like the interaction potential energy, in a bottom-up manner from simulations with the fine-grained model. This approach, which is based on the Mori-Zwanzig projection operator formalism, has been successfully utilized in several recent publications to parameterize coarse-grained models that consistently model the dynamics of model systems at low density. However, the approach relies on assumptions on the nature of the system which are not fully satisfied at the higher density typical for soft matter systems. Most importantly, it is assumed that the degrees of freedom removed from the model upon coarse-graining relax infinitely fast in comparison with those retained in the coarse-grained model (complete time scale separation). In this thesis, an application of such a procedure for a dynamically consistent coarse-grained model is presented for realistic model systems of soft matter. The aim of this study is to evaluate whether Mori-Zwanzig-based coarse-grained models can be used for the simulation of

realistic soft matter systems, in which time scale separation is not complete. To this end, coarse-grained models are parameterized for model systems of different chain length and the predictions of the dynamics produced by the coarse-grained system are compared to those of an atomistic reference model. The self-diffusion coefficients of these systems can be reproduced to a good degree, whereas dynamics properties with a smaller characteristic time scale are less well reproduced with the Mori-Zwanzig coarse-grained model, a finding that can be related to the increasing deviation of the system's state from the assumed complete time scale separation.

Zusammenfassung in deutscher Sprache

Computergestützte Simulationen von Molekülen und Atomen sind von erheblichem Wert für die Erforschung der weichen Materie. Die physikalische Chemie profitiert in großem Maße von den Erkenntnissen, die mithilfe von klassischen molekulardynamischen Simulationen gewonnen werden können. In solchen Simulationen werden Atome als Punktmassen dargestellt, deren Bewegungen den Newtonschen Gesetzen folgen. Eine große Herausforderung bei der Druchführung solcher Simulationen liegt in der Tatsache begründet, dass eine Simulation von vergleichsweise kleinen Systeme mit atomistischer Auflösung nur unter Einsatz großer Rechenleistung durchzuführen sind und die Zeit- und Längenskalen, die mit atomistischen Simulationen abgebildet werden können, deshalb stark begrenzt sind.

Ein Ansatz, Simulationen eines gegebenen Systems effizienter zu gestalten ist die Vergröberung des Modells (engl. "coarse-graining"), d.h. die Reduktion der räumlichen Auflösung durch Vereinigung einzelner Atome zu größeren Modelleinheiten. Coarse-graining geschieht in zwei Schritten: zunächst muss eine Zuordnung der Modelleinheiten zwischen den einzelnen Skalen definiert werden (das sog. "mapping"), im zweiten Schritt müssen Wechselwirkungen zwischen den Modelleinheiten des gröberen Modells definiert, bzw. berechnet werden.

Besonders der zweite Schritt ist anspruchsvoll, wenn der Anspruch an das vergröberte Modell, möglichst viele Eigenschaften des zugrundeliegenden Referenzmodells korrekt abzubilden, erfüllt werden soll. In der Literatur existieren einige Methoden hierzu, die einer gemeinsamen Grundidee, dem systematischen, bottom-up coarse-graining, folgen. In diesen Methoden werden die Wechselwirkungen der vergröberten Auflösung aus den (meist atomistischen) Wechselwirkungen des Referenzmodells berechnet.

Eine wichtige Qualität der vergröberten Modelle im Allgemeinen ist die sog. Transferierbarkeit, d.h. die Fähigkeit die Eigenschaften des zugrundeliegenden Referenzmodells auch an Zustandspunkten, an denen das vergröberte Modell nicht parametrisiert wurde, korrekt wiederzugeben. Unter den existierenden coarse-graining Methoden ist insbesondere die "Conditional Reversible Work" (CRW) Methode in der Lage, transferierbare vergröberte Modelle zu parametrisieren. Die CRW Methode ist hierbei sowohl konzeptuell als auch in der Implementierung einfach und effizient in Bezug auf die benötigte Rechenleistung.

In dieser Dissertation werden Studien präsentiert, welche die Erweiterung der Anwendungen von CRW-basierten Modellen zum Ziel haben. Die CRW-Methode wird, neben anderen, in einer komparativen Studie genutzt, um vergröberte Modelle zu entwickeln, die die Dampf-Flüssig-Gleichgewichte und Mischungsthermodynamik von Hexan und Perfluorohexan korrekt beschreiben. Die Ergebnisse dieser Studie zeigen, dass die Transferierbarkeit eines Modells stark von der Methode abhängt, mit der die Wechselwirkungspotentiale parametrisiert werden. Weiterhin wird gezeigt, dass die CRW-basierten Modelle eine gute Transferierbarkeit in Bezug auf die Temperatur, den Transfer eines Moleküls von der Gas- in die Flüssigphase und der Veränderung der Zusammensetzung einer binären Mischung besitzen.

Eine weitere Neuerung in der CRW-Methodik, die in dieser Dissertation vorgestell wird, ist die Erweiterung der Methode für die Parametrisierung von Wechselwirkungen in polaren Systemen. Bis dato, sind nur CRW Modelle von apolaren Systemen veröffentlicht worden. Hier wird gezeigt, dass eine leicht modifizierte Methode, die separate Parametrisierung von elektrostatischen und Van-der-Waals Wechselwirkungen ermöglicht. Die resultierenden Modelle für schwach polare organische Verbindungen und ionische Flüssigkeiten zeigen im Wesentlichen die gleichen Qualitäten, wie die CRW Modelle von apolaren Systemen.

Die konsistente Darstellung von dynamischen Eigenschaften in vergröberten Modellen ist eine weitere große methodische Herausforderung. Im Allgemeinen wird die Dynamik eines Systems von vergröberten Modellen schneller dargestellt, als von dem zugrundeliegenden hochaufgelöstem Modell. Der Grund für dieses Verhalten ist die Reduktion von Energiebarrieren im vergröberten Modell. Es ist möglich, die

Auswirkungen dieses Effekts durch nachträgliches Einführen von Reibungstermen in die molekularen Wechselwirkungen zu kompensieren. Diese Reibungsterme können, genau wie die konservativen Teile der Wechselwirkungen, in einem bottom-up Verfahren, welches auf den Projektionsoperatoren nach Mori und Zwanzig basiert, aus den Wechselwirkungen im hochaufgelösten Modell bestimmt werden. Solche Modelle wurden in der Literatur bereits für einige Modellsysteme erfolgreich angewendet. Um leistungsfähige Mori-Zwanzig-Modelle für vergröberte Systeme zu parametrisieren sind allerdings Annahmen vonnöten, welche für komplexe, molekulare Systeme mit hoher Dichte im Allgemeinen nicht erfüllt sind. Am stärksten fällt hierbei die Annahme ins Gewicht, dass die Zeitskalen der Freiheitsgrade, die während des coarse-grainings aus dem Modell entfernt werden, unendlich schnell gegenüber jenen, welche im vergröberten Modell verbleiben, relaxieren (vollständige Trennung der Zeitskalen). In dieser Dissertation wir eine Anwendung des Mori-Zwanzig-basierten coarse-grainings für realistische Systeme weicher Materie präsentiert. Die Fragestellung dieser Studie ist, inwiefern bzw. zu welchen Grad eine konsistente Reproduktion dynamischer Eigenschaften für solch ein System, in dem die komplette Zeitskalentrennung nicht gegeben ist, möglich ist. Hierzu werden vergröberte Modelle für molekulare Modellsysteme mit verschiedenen Kettenlängen parametrisiert. Ausgewählte dynamische Eigenschaften dieser Modelle werden anschließend mit dem hochaufgelösten Modell verglichen. Hierbei zeigt sich, dass die Eigendiffusionskoeffizienten der Modellsysteme gut reproduziert werden. Andere Eigenschaften, welche eine kleinere charakteristische Relaxationszeit aufweisen, werden von den vergröberten Modellen schlechter reproduziert, wobei die Abweichung von der hochaufgelösten Referenz mit steigendem Polymerisationsgrad größer wird; ein Verhalten, dass durch die zunehmende Abweichung des Systems von der vollständigen Zeitskalentrennung erklärt werden kann.

Acknowledgements

A long time at the TU now comes to a close. This is the point where I would like to say "Thank you" to a the people who accompanied me on the way.

Special thanks go to my supervisor Prof. Dr. Nico van der Vegt for the excellent support during my time as a PhD candidate. His enthusiasm for science, the many inspiring discussions, and the freedom to pursue my own scientific ideas independently have contributed to a large degree to the success of my research projects.

I also thank the second referee to this thesis Prof. Dr. Müller-Plathe for the time taken to grade my work, and for the interesting discussions in the countless theory seminars of the last years. Further, I would like to thank the other members of my examination board, PD Dr. Immel and Prof. Dr. Krewald, for their effort in the examination of my thesis.

I would like to express my gratitude to the Deutsche Forschungsgemeinschaft for the generous funding of this thesis through the TRR-146 "Multiscale Simulation Methods for Soft Matter Systems".

A big "Thank you" to my coworkers in the Van-der-Vegt group for the great collaboration. Thanks Marvin and Viktor, David, Valentina, Emiliano, Claudia, Francisco, Imke, Tobias, Severine and all the others.

The many years in Darmstadt would not have been the same without my friends. So thanks a lot to Fabian, Henner, Günni, Marvin, Kai, Velten, Aki, Jonas, and to all the others who accompanied me on my way to this point.

Last, but certainly not least, I owe the most sincere gratitude to my family. Thank you Judith and Johanna for enduring all my highs and lows during my work on this thesis and thank you for giving me the pats on the back I sometimes needed to keep going on. A big "Thank you" also to Mama, Papa, and Eva for your support during all my studies.

Contents

1 Introduction

Computers have transformed daily life to a degree and at a rate which is without precedent in human history. For decades before computers became widespread as household appliances, computer simulations have been utilized in the physical sciences. From the very beginning, one of the focuses of computer simulations was the research of problems in physical chemistry and statistical mechanics by the simulation of ensembles of particles in motion.[1–4]

Simulations in these fields have, of course, profited immensely from the increase in computational performance within the last thirty years. While initial particle-based simulations contained only a few dozens of Lennard-Jones-type model atoms[2], atomistic simulations of e.g. large biomolecules in aqueous solution or high-molecular-weight macromolecular systems are possible today[5, 6]. Still, not every system and problem of interest can be studied with models at an atomistic scale.

Simulations with these highly-resolved models on macroscopic time and length scales are impossible with existing computational resources, and even if it were possible to simulate every process at atomistic (or even electronic) resolution, it would certainly not be wise to do so indiscriminately. The reason for this is simple: physical effects and processes occur on characteristic time and length scales and can be studied efficiently with models that share these scales, so the choice of the ideal model resolution depends on the problem the researcher is interested in. Complex physical processes in the real world often include causes and effects on different scales and a good simulation of such processes has to span over these scales to faithfully predict their outcome. Simulations for complex phenomena therefore feature different sub-scales with sub-problems at different length and time scales and feed their predictions back into simulations of the larger and/or smaller length scale. This simulation paradigm is called scale-bridging, hierarchical, or multi-scale modeling.[7] The scale transition which is arguably most important for the research of physical chemistry of soft matter is the transition from the atomistic resolution to larger time and length scales.[7, 8]

The process of reducing the model resolution is referred to as coarse-graining and the resulting models are called coarse-grained (CG) models. The term CG model describes a wide stretch of model resolutions on the nano-scale, where one CG interaction site may represent any number between two to several thousands of atoms. In the context of this work, a CG model is a model in which low numbers (ca. 2-20) of atoms are merged into 'super-atoms'. The interaction potentials of models with this degree of coarse-graining resemble atomistic interactions and it is possible to retain a considerable degree of chemical detail in these CG models. These characteristics make CG models of this variant useful in the research of the physics and physical chemistry of liquids (molecular and ionic), (bio-)polymers in bulk or solution, and their behavior at interfaces.[7–10]

The interactions sites of a CG model represent material-specific characteristics through a set of potential energy terms, which are functions of the spatial configuration of the sites. Determining these potential energy terms can be achieved by the utilization of information from an underlying fine-grained (FG) model (which is usually atomistically resolved). A considerable amount of research has been performed on the algorithms required for the extraction of CG interaction parameters from FG simulations (bottom-up or systematic coarse-graining). The properties and qualities of a CG model depend significantly on the algorithm chosen for their parameterization.[8, 10]

One of the most challenging tasks in the development of a CG model is to parameterize interactions between the CG particles such that the model predicts correct thermodynamics and structure not only at the state point of parameterization but also different state points.[11] This quality, the transferability of the model, has been of interest for a long time and the challenge of parameterizing transferable CG models has been tackled in the past using a variety of different coarse-graining methods and a-posteriori modifications of CG models.[12–23]

The conditional reversible work (CRW) and effective-force coarse-graining (EF-CG) methods have been developed specifically with the intent of parameterizing transferable CG models for soft matter systems.[24, 25] In these methods, which are conceptually very similar, the CG interaction potential is calculated from pair forces between the atoms constituting the "superatom" represented by the CG interaction site. They can therefore be referred to as "pair-interaction-based" coarse-graining methods.

In past studies, the good transferability of CRW models has been demonstrated for multiple systems using aliphatic and aromatic hydrocarbons as model systems.[24–30] One of the aims of this thesis is to study the benefits and range of applicability of the pair-interaction-based coarse-graining approach. To this end, studies on the calculation of thermodynamic properties that have not previously been studied with these models, as well as the parameterization of new CRW models for systems beyond the apolar hydrocarbons have been undertaken.

In chapter 3 of this thesis an additional case study to demonstrate the transferability of CRW and EF-CG models for apolar liquids is presented. To test the performance of different CG models with respect to the state point transferability the vapor-liquid and liquid-liquid phase equilibria of hexane and perfluorohexane are studied. Simulations with the CRW and EF-CG models reproduce the vapor-liquid coexistence curve of a united-atom reference to a good degree. Further, these models capture the enthalpy-entropy decomposition of the free enthalpy of vaporization well, even for temperatures far from the temperature of parameterization. While a reproduction of the liquid-liquid phase separation could not be observed in simulations, it is shown that the CRW and EF-CG models are transferable with respect to the composition of a liquid mixture and reproduce the mixing enthalpies and volumes better than an unrefined CG model parameterized using structural information from the FG reference system. These new results illustrate a possible new application for CRW and EF-CG models and further demonstrate the capability of the CRW method to parameterize transferable CG models in a computationally inexpensive manner.

Like in older studies on CRW modeling, the model compounds used chapter 3 are apolar molecular liquids. These systems are comparatively easy to model as the long-ranged electrostatic interactions only play a minor role in the behavior of the system. However, many interesting soft matter systems consist of polar and ionic compounds, and a coarse-graining algorithm intended for a general use in the simulation of soft matter therefore needs to be able to model the behavior of such systems adequately. In chapters 4 and 5 of this thesis, a modification of the CRW method is presented which is introduced to enhance the modeling with mappings, that are defined such that the CG interaction sites present a non-zero charge. In the modified CRW scheme, two pair interaction potentials are obtained: one representing the effective electrostatic and another representing the effective Van-der-Waals interactions. Further, the electrostatic component can, in many cases, be approximated well by a Coulomb potential that uses the sum of partial atomistic force field charges on the CG interaction site as an effective site charge. The model systems used in the two chapters are two non-ionic compounds with polar functional groups and a series of butylimidazolium ionic liquids in chapters two and three, respectively. The results in both cases are similar: the transferability observed for nonpolar systems is also observed for the studied polar model compounds, especially the transferability with respect to temperature and the correct reproduction of the liquid-vapor interfacial tension.

In summary, the first three chapters serve as a strong evidence that the coarse-graining methods based on direct pair interactions provide a conceptually intuitive, computationally efficient way to obtain CG models that include a good degree of transferability.

Besides the question of transferability, another important challenge of CG multi-scale modeling in soft matter is the consistency of dynamic properties between model scales.[31–35] Coarse-graining flattens the model's free energy surface, with the result that dynamics are faster in CG models than in an FG representation of the same system. When only thermodynamic quantities are of interest, this effect is not problematic, rather the acceleration of relaxation dynamics contributes in this case to the computational efficiency of the model, and in certain cases the acceleration can be corrected for by the use of scaling factors.[35] However, when dynamic quantities or processes which are kinetically rather than thermody-

namically driven are to be studied, CG models that represent the dynamics of the FG scale consistently become essential.[35, 36]

The existence of a speed-up upon coarse-graining and its causes are a known problem in statistical physics and an approach exists to model coarser model scales in a manner which is dynamically consistent with the underlying FG reference representation. This approach, the projection operator (PO) formalism, uses the known description of the time-evolution of the FG phase space coordinates to derive an equation for the time-evolution of the CG coordinates.[37–39] In addition to a consistent description of the CG interactions, such a model also includes a CG equation of motion (EOM) which differs from the Newtonian EOM utilized for the time-propagation of the FG model. In addition to the conservative interactions originating from the CG interaction potentials, the CG-EOM contains a drag force which explicitly depends on previous states of the model and is therefore referred to as memory term, and a random force which is related to the memory term through a fluctuation-dissipation theorem.[37–39] The additional contributions to the CG-EOM do not influence the ensemble averages of static properties of the system, but will reduce the relaxation times of the CG degrees of freedom (and hence 'slow-down' the accelerated CG dynamics).

In CG systems with Markovian behavior, i.e. systems where the relaxation of the 'fast' degrees of freedom (DoF) (i.e., the DoF which are lost upon coarse-graining) is infinitely fast in comparison with the 'slow' DoF (which are retained in the CG model) the explicitly time-dependent memory term can be approximated by an instantaneous dissipative interaction between CG interaction sites.[40] Using the Markovian approximation, which is analogous to the Born-Oppenheimer approximation used in quantum mechanics, the EOM of the CG system is equivalent to the EOM utilized in the dissipative particle dynamics (DPD) simulation method[34, 40, 41], which is a well-studied method for the simulation of models at coarse scales.[42–45]

The DPD EOM adds to the conservative force originating from the interaction potentials drag and noise terms, which are both dependent on distance-dependent friction coefficients. Like the conservative CG interactions, these friction coefficients can be parameterized in a bottom-up manner from dynamic information of a FG reference system. The resulting models are also referred to as Mori-Zwanzig DPD (MZ-DPD) models, after the two authors which are usually credited with the formulation of the PO formalism. In recent years, a considerable number of studies have been published studying the addition of DPD friction parameterization to existing coarse-graining methods.[34, 46–56] Almost all of the models presented in these studies utilize mappings in which one (or more) molecules are contained completely within one CG interaction site, and in which the particle number density is often comparatively low. This model design is often chosen because it leads to a Markovian character of the model system, which in turn makes an accurate approximation of the CG dynamics with the DPD-EOM approach possible. However, systems with these characteristics are untypical in CG models of soft matter, where the polymeric nature of many studied systems dictates that there be multiple sites per molecule. Such an arrangement of the CG model will be described in this thesis as a "molecular mapping". For systems at realistic density and with a molecular mapping, the Markovian assumption is usually not valid. This thesis addresses the question whether it is nonetheless possible to reproduce the fine-grained dynamics with MZ-DPD models.

In chapter 6, the parameterization of a bottom-up DPD model and the properties of the resulting CG model are studied using molecules with one, two and 24 sites per CG molecule. The dynamics of liquids modeled with molecular mappings can be reproduced on diffusive time scales, while the MZ-DPD models lack representability with respect to the dynamic behavior on shorter time scales, especially in reproducing structural and velocity time-correlation functions quantitatively. A novelty presented in this chapter is the simulation of polymer permeation with a MZ-DPD model. This process is of particular interest because its speed-up depends in a nontrivial manner on the change of the altered free energy surface of the CG model.[36] Results indicate that adding dissipative interactions to the CG model is not sufficient for the consistent modeling of the diffusion of small penetrant molecules in a polymer matrix.

2 Theoretic Background

2.1 Molecular Modeling of Soft Matter Systems

Since the advent of mechanized computing, numerical simulations of the physical world have been performed. Now, with powerful computers being and available at low expense, simulations have become a standard research technique for the study of problems in science and engineering. The simulation is now understood as a third pillar of science, along with theory and experiment. Occasionally the term "in-silico experiment" is used to refer to computer simulations and this term describes well the intent of the scientific study: the reproducible, iterative application of a set of physical laws that are axiomatically imposed onto the system. These laws can be completely identical to those governing the real world (and hence an experiment in the laboratory) but they can also differ from those. A good "simulation universe" is designed such that the most important physical phenomena for the problem at hand are reproduced and insignificant detail is omitted for the sake of increased computational performance; in this trade-off lies one of the challenges to the simulator.[3, 4, 57]

A typical simulation study with particle-based models consists of two steps. First, an ensemble of states is generated in a deterministic or stochastic manner for the Hamiltonian of the system. Then this ensemble of states is analyzed, using statistical physics, to calculate the desired properties of the system. Simulations of this type are now omnipresent in the physical sciences and cover, among others, the research of quantum physics and chemistry, solid state systems, polymers and biological systems.[3, 4, 57]

This thesis presents methodological studies on molecular simulation techniques for the use in soft matter research. The term 'soft matter' denotes a wide range of condensed phase materials, in which the energy barriers are of the order of the thermal energy at ambient temperature. Following this definition almost any material which is neither cristalline solid nor gaseous can be characterised as soft matter. For the description of the physical chemistry of such systems at near-ambient conditions (assuming the absence of chemical reactions) non-relativistic, classical models are usually sufficient in which the conserved energy takes the form of a Hamitonian $\mathscr{H}$:

$$\mathscr{H} = U(\mathbf{r}) + T(\mathbf{p}), \tag{2.1}$$

where U and T are potential and kinetic energies, respectively, which depend on a set of coordinates $\mathbf{r}$ and the respective set of conjugate momenta $\mathbf{p}$. The model coordinates can, in principle, be chosen arbitrarily, however, the $\mathbf{r}$ are often chosen as real-space coordinates of N particles described as point masses, for practical reasons. In this case the Hamiltonian can be simplified to read:

$$\mathscr{H} = U(\mathbf{r}) + \sum_{i}^{N} \frac{\mathbf{p}_i^2}{m_i}, \tag{2.2}$$

using the masses m_i and translational momenta $\mathbf{p}_i$ of the ith particle.

The creation of a model for a system at hand is now reduced to the choice of the proper set of model coordinates and a corresponding term for the potential energy of the system. In atomistic models positions of individual atoms are chosen as model coordinates. Other representations of materials are also widely used and these may use functional groups[58] or molecular centers of mass[7, 8, 10] or charge[59] as basic sites. The choice of resolution is closely related to the purpose of the model. Lowering the resolution of a model will reduce the computational expense of simulations whereas increasing the resolution will provide more insight into microscopic behavior and allows for the inclusion of more chemical detail into the model.

The second requirement for a complete molecular model is the appropriate choice of the potential energy function $U(\mathbf{r})$ which is typically formulated in terms of interaction energies between the sites of the model. For molecular models, $U(\mathbf{r})$ will have two types of contributions: a bonded and nonbonded part:[7]

$$U(\mathbf{r}) = U_{\text{nonbonded}}(\mathbf{r}) + U_{\text{bonded}}(\mathbf{r}_{\text{bond}}, \boldsymbol{\theta}, \boldsymbol{\phi}) \tag{2.3}$$

The nonbonded interactions act between all sites which are not close neighbors in the same molecule. This contribution usually energetically penalizes close vicinity of sites, this is a prerequisite to model collisions. Further, it may contain attractive interactions to model a cohesive character of the system. In atomistic models, the nonbonded potential is in most cases a combination of the Coulomb potential and the Lennard-Jones potential representing electrostatic and induced dipole interactions, respectively:

$$U_{\text{nonbonded}}{}^{ij}(r_{ij}) = 4\varepsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) + \frac{q_i q_j}{4\pi\varepsilon_0}\frac{1}{r_{ij}}, \tag{2.4}$$

where σ_{ij} and ε_{ij} are parameters specific to the pair of atoms i and j, q_i and q_j describe the partial charges of these atoms, and ε_0 is the dielectric permittivity of the vacuum. In most models, the nonbonded interactions are formulated on a pairwise basis. The bonded interactions, on the other hand, serve as an energetic penalty term which is placed onto the system to conserve the desired molecular geometry. These interactions also include pairwise terms but often, especially in atomistic systems, the complex and rigid geometry of molecules requires the use of higher-order interactions. More specifically, three- and four-body interactions can be used to model the correct bond and torsion angles, denoted above as θ and ϕ, respectively, within a molecule. The functional form and the parameters of these potential functions can be determined from theoretical considerations, experimental data, or simulations at lower resolution of the electronic state of a single molecule.

2.2 Molecular Simulation Paradigms

The aim of molecular simulation algorithms is to generate a series of configurations $\mathbf{r}$, the trajectory, from which thermodynamical quantities can be calculated using principles of statistical physics. The relative weight of a configuration is proportional to its potential energy U through the Boltzmann factor $\exp(-\beta U)$, where $\beta = (k_{\text{B}}T)^{-1}$ is the inverse of the (absolute) temperature T multiplied with the Boltzmann constant k_{B}.[57] Two main paradigms have evolved for the efficient numerical generation of trajectories in computational soft matter research: molecular dynamics (MD)[2] and Metropolis-Criterion Monte Carlo (MC)[1]. These paradigms are fundamentally different from one another and each has strengths and weaknesses that render it suitable for the application to different problems and systems.

2.2.1 Metropolis-Criterion-Monte-Carlo

In MC simulations the sampling is performed by generating new configurations using random numbers. Trial configurations are generated from the previous configuration using so-called moves, which are randomly generated displacements of particle coordinates. These moves may be, but are not limited to, translations and rotations of molecules, modification of molecule geometry, or deleting and reinserting molecules from the simulation space. After generating the trial configuration its potential energy is compared to the previous configuration yielding $\Delta U = U_{trial} - U_{prev}$. The trial configuration is accepted as the new configuration of the system with the following probability p:[4]

$$p(\Delta U) = \begin{cases} 1 & \text{if } \Delta U < 0 \\ e^{-\beta \Delta U} & \text{if } \Delta U \geq 0 \end{cases}. \tag{2.5}$$

The idea behind this approach is that the system can be equilibrated fast from phase space regions of high potential energy, while allowing for moves that increase the potential energy within the thermal fluctuations. The resulting trajectory is Boltzmann-distributed after the system is equilibrated.

An advantage of MC sampling lies in the possibility to generate unphysical types of moves which can accelerate the rate of phase space sampling.[3] Also, the system can explore regions in the phase space that are not accessible to physical motion in phase space due to high activation energy barriers. MC simulation is disadvantageous in systems that require a large number of trials to obtain an accepted configuration and systems where concerted motion of different entities is required to generate new trial configurations. Systems that are best studied with MC simulation are therefore relatively simple and show a low density or high temperature. Prime examples of such systems are gases and supercritical fluids. Especially for liquid systems at relatively low temperature (these are the characteristics of most soft matter applications) MC simulations tend to be inefficient. Another disadvantage of MC simulations is that no information whatsoever is generated regarding the velocities of the simulated entities. Therefore no quantities that are related to the dynamics or the kinetic energy distribution (i.e. the instantaneous temperature) of the system can be calculated from MC-generated trajectories.[3]

As a consequence of these disadvantages of the MC paradigm, molecular dynamics simulations are often better suited to investigate soft matter problems. In MD simulations the trajectory is generated in a physical manner and includes information on the dynamics of the system.

2.3 Molecular Dynamics

Molecular dynamics is a simulation scheme that numerically solves Newton's equation of motion for a system of atoms. The system can be described using the Hamiltonian $\mathscr{H}$ which can be expressed as the sum of potential energy U and kinetic energy T. In terms of the atomistic coordinates ($\mathbf{r}$), momenta ($\mathbf{p}$), and masses (m) the Hamiltonian is defined in the well-known manner:

$$\mathscr{H} = \sum_i \frac{|\mathbf{p}_i|^2}{2m_i} + U(\mathbf{r}). \tag{2.6}$$

2.3.1 Equation of Motion in MD

In a system with given initial positions $\mathbf{r}$ and momenta $\mathbf{p}$ the Newtonian equations of motion for the ith atom read:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} \tag{2.7}$$

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial U}{\partial \mathbf{r}_i} = \mathbf{F}_i, \tag{2.8}$$

where $\mathbf{F}_i$ is the force on the ith atom. Using these equations, a numeric iterative scheme can be derived to propagate the positions and momenta in time.

The resulting trajectory contains all information on the positions, momenta of and forces on all the atoms in the system. The configurations within the trajectory are Boltzmann-distributed if the system is in equilibrium. In addition to the thermodynamic properties emerging from the positions of the atoms, MD-generated trajectories can be used to calculate dynamic quantities and to study temperature fluctuations within the system.

2.3.2 Thermostats in MD Simulation

The MD simulation algorithm described so far will generate a NVE-ensemble with constant atom number, volume and energy. For reasons discussed below, it is often more desirable to sample the NVT-ensemble with constant temperature, rather than constant energy. To achieve this sampling, the simulation scheme has to be adapted. The motion of the atoms of the system has to be coupled to a heat bath that scales the velocities of the atoms such that the desired temperature is reached. The microscopic temperature of the system is defined through the atom velocities $\mathbf{v}$ by:

$$\sum_i m_i |\mathbf{v}_i|^2 = N_{\mathrm{DoF}} k_{\mathrm{B}} T, \tag{2.9}$$

where N_{DoF} is the number of thermal degrees of freedom of the system (in most simulations $N_{\mathrm{DoF}} \approx 3N$)

Average quantities are in principle the same as those calculated from a trajectory sampled in the NVT-ensemble (assuming a system in equilibrium, of course), but the simulation of NVT-ensembles has certain benefits over NVE simulations. First, the equilibration of a newly set-up configuration of atoms can only be meaningfully performed in the NVT-ensemble, where the heat bath can act to enforce the correct velocity distribution onto the atoms.[3, 4] Second, numerical imprecision leads to a drift in the total energy of NVE-simulations on longer time scales. This drifting energy is absorbed by the heat bath and thereby an accumulation of numerical errors is prevented. A few of the most important thermostat algorithms will be introduced in the following.[3, 4]

Velocity rescaling and weak-coupling

This class of thermostat algorithms works ensures the correct system temperature by scaling all velocities of the system by a constant factor λ:

$$\mathbf{v}_{new} = \lambda \mathbf{v}_{old} \tag{2.10}$$

The scaling factor for the velocities is calculated by comparing the instantaneous temperature to the reference temparature of the heat bath T_{ref}.[60] For simple velocity scaling:

$$\lambda = \sqrt{\frac{T_{\mathrm{ref}}}{T}}. \tag{2.11}$$

This can lead to strong modifications of the velocities in the system (if $\lambda \not\approx 1$) and therefore leads to a massive perturbation of the dynamics of the system. A remedy to this problematic behavior of the simple velocity rescaling thermostat is contained in the so-called weak-coupling temperature coupling (also referred to as Berendsen thermostat).[61] Here the scaling in velocity is not instantaneous but gradual, by modifying the calculation of λ:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_{\mathrm{T}}}\left(\frac{T_{\mathrm{ref}}}{T} - 1\right)\right]^{\frac{1}{2}}, \tag{2.12}$$

where Δt is the time step used to propagate the simulation. Here a time scale τ_{T} is introduced on which the temperature deviation from the reference value is decaying. The decay of the temperature difference is a first order process, and this leads to a smoother change in velocities that does not perturb the dynamics as drastically as the instantaneous rescaling. The weak-coupling algorithm is very robust when used for systems far from equilibrium. It is therefore widely used for equilibration purposes. For production simulations, it is less suitable because it does not create a physical ensemble sampling, which leads to erroneous fluctuations of thermodynamic quantities.[4]

Nosé-Hoover chains

For production simulations in which the correct fluctuations of thermodynamic quantities is important, thermostat algorithms that go beyond rescaling atom velocities have to be employed, such as the Nosé-Hoover thermostat algorithm.[62, 63] This algorithm extents the system's Hamiltonian by a heat bath degree of freedom ζ and a corresponding conjugated momentum p_ζ:

$$\mathscr{H} = \sum_i \frac{|\mathbf{p}_i|^2}{2m_i} + U(\mathbf{r}) + \frac{p_\zeta^2}{2m_\zeta} + N_{\mathrm{DoF}} k_{\mathrm{B}} T \zeta. \tag{2.13}$$

The equation of motion changes accordingly:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_\zeta}{m_\zeta}\mathbf{v}_i, \tag{2.14}$$

and an equation of motion for ζ is defined using the differences in instantaneous and reference temperature:

$$\frac{dp_\zeta}{dt} = (T - T_{\mathrm{ref}}). \tag{2.15}$$

In the definitions above, a conjugated mass of the heat bath coordinate m_ζ was introduced. This heat bath math can be related to the reference temperature and coupling time constant:

$$m_\zeta = \frac{\tau_{\mathrm{T}}^2 T_{\mathrm{ref}}}{4\pi^2} \tag{2.16}$$

The simple Nosé-Hoover-thermostat suffers from ergodicity problems, but these problems can be overcome by coupling ζ to a second Nosé-Hoover-thermostat, which is in turn coupled to third heat bath, and so on. This algorithm is called a Nosé-Hoover chain of length N, where N is the total number of thermostat levels.[3]

Langevin thermostat

Another way of modifying the equation of motion in order to generate a NVT-ensemble is by using the Langevin equation of motion for the simulation.[64] This equation of motion adds a dissipative and random force (F_{D} and F_{R}) to the force of each particle. The modified equation of motion reads:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i + \mathbf{F}_{\mathrm{D},i} + \mathbf{F}_{\mathrm{R},i} \tag{2.17}$$

$$= \mathbf{F}_i - \frac{m_i}{\tau_{\mathrm{T}}}\mathbf{v}_i + \sqrt{\frac{2m_i k_{\mathrm{B}} T_{\mathrm{ref}}}{\tau_{\mathrm{T}}}}\xi_i, \tag{2.18}$$

where ξ_i is a vector of Gaussian random numbers with zero mean and unit variance. In a physical sense, this algorithm creates random collisions with a heat bath of virtual particles with temperature T_{ref}. The Langevin thermostat does not require the calculation of the instantaneous temperature to generate a correct NVT-ensemble. Therefore it is well suited for the use in systems with a small number of particles, where the natural fluctuation of the instantaneous temperature is rather large. The thermostat algorithms introduced above will, in this case, become unstable or lead to strong modifications of the dynamics of the system. A drawback of the Langevin thermostat is that the coupling time constant has to be carefully chosen. If it is too large, the thermostat will not modify the dynamics strong enough to exert any effect on the temperature. The simulation then effectively becomes a NVE-simulation. In the limit of a too small τ_{R} the dynamics of the system will be strongly dampened. In this case a NVT-ensemble is still simulated, but the dynamics is completely different to the dynamics that would have been obtained from a simulation under NVE conditions or with the use of an instantaneous-temperature-based thermostat.

Dissipative particle dynamics thermostat

Dissipative particle dynamics (DPD) was initially developed as a simulation algorithm for large and soft particles (e.g. colloids in solution). [42–44] The DPD equation of motion is similar to the Langevin equation of motion: it contains the same three contributions to the total force. The major difference lies in the fact, that the dissipative and random force are calculated from pairwise contributions. These pair-friction forces depend on a distance-dependent pair friction function $\gamma_{ij}^{\parallel}$. The DPD equation of motion reads:

$$\begin{aligned} m_i \frac{d^2\mathbf{r}_i}{dt^2} &= \mathbf{F}_i + \mathbf{F}_{\mathrm{D},i} + \mathbf{F}_{\mathrm{R},i} \\ &= \mathbf{F}_i - \sum_j \gamma_{ij}^{\parallel}(r_{ij})\mathbf{v}_{ij}^{\parallel} + \sum_j \sqrt{2\gamma_{ij}^{\parallel}(r_{ij})k_{\mathrm{B}}T_{\mathrm{ref}}}\boldsymbol{\xi}_{ij}, \end{aligned} \tag{2.19}$$

and uses the projection of the velocity difference on the ij-distance vector:

$$\mathbf{v}_{ij}^{\parallel} = \left(\mathbf{v}_{ij}\cdot\mathbf{e}_{ij}\right)\mathbf{e}_{ij} \tag{2.20}$$

$$\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j \tag{2.21}$$

$$\mathbf{e}_{ij} = \mathbf{r}_{ij}/|\mathbf{r}_{ij}|. \tag{2.22}$$

As a consequence of this modification to the Langevin EOM, the friction coefficient acting on a particle is dependent on the surroundings. In this manner, collisions between soft particles can be realistically described. In conjunction with soft interaction potentials, a quadratically decaying functional form of $\gamma_{ij}^{\parallel}$ is often used in traditional DPD simulations:

$$\gamma_{ij}^{\parallel} = \gamma_{0,ij}\left(1 - \frac{r_{ij}}{r_{\mathrm{C}}}\right)^2, \tag{2.23}$$

where r_{C} is a cutoff length and $\gamma_{0,ij}$ is a constant parameter. The superscript $\parallel$ is used here to underline the fact that the dissipative and random forces related to $\gamma_{ij}^{\parallel}$ are parallel to the ij-distance vector.

Using this simulation method it is possible to parametrize models for soft matter systems that are capable of reproducing hydrodynamic properties for soft matter systems at a rather low degree of resolution. The DPD thermostat, in its original formulation, conserves the total momentum of the system, which is an important precondition for the correct prediction of hydrodynamic behavior of a system. It has been observed, however, that for systems at higher density the standard DPD approach will not be able to reproduce experimental shear viscosity. An attempt to cure this mismatch has been formulated by introducing a second (transversal) friction term into the equation of motion.[65] The forces related to this friction depend on the velocity component perpendicular to the ij-distance vector $\mathbf{v}_{ij}^{\perp} = \mathbf{v}_{ij} - \mathbf{v}_{ij}^{\parallel}$. A second friction function $\gamma_{ij}^{\perp}$ is introduced which is independent of the parallel friction function. The equation of motion of this extended DPD method reads:

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i + \mathbf{F}_{\mathrm{D},i}^{\parallel} + \mathbf{F}_{\mathrm{R},i}^{\parallel} + \mathbf{F}_{\mathrm{D},i}^{\perp} + \mathbf{F}_{\mathrm{R},i}^{\perp} \tag{2.24}$$

$$= \mathbf{F}_i - \sum_j \gamma_{ij}^{\parallel}(r_{ij})\mathbf{v}_{ij}^{\parallel} + \sum_j \sqrt{2\gamma_{ij}^{\parallel}(r_{ij})k_{\mathrm{B}}T_{\mathrm{ref}}}\boldsymbol{\xi}_{ij}^{\parallel} \tag{2.25}$$

$$- \sum_j \gamma_{ij}^{\perp}(r_{ij})\mathbf{v}_{ij}^{\perp} + \sum_j \sqrt{2\gamma_{ij}^{\perp}(r_{ij})k_{\mathrm{B}}T_{\mathrm{ref}}}\boldsymbol{\xi}_{ij}^{\perp} \tag{2.26}$$

2.3.3 Dynamical Quantities from MD Simulations

In addition to static average quantities, MD simulations also allow for the study of dynamic processes and this fact can be used to gain study properties relating to spectroscopy, rheology, or heat-transfer in soft matter systems.

For the calculation of such properties, a time correlation function C is used which follows the general form:

$$C(\tau) = \langle x(0) \cdot y(\tau) \rangle, \tag{2.27}$$

where x and y is a properties extracted or calculated from the trajectory data. In the special case where x and y describe the same property the correlation function is called auto-correlation function (ACF).

Important properties of the ACF are:

$$C(0) = \langle x^2 \rangle \tag{2.28}$$

$$C(\infty) = \langle x \rangle^2 \tag{2.29}$$

In the special case $\langle x \rangle = 0$ this means that $C(0) = \mathrm{var}(x)$ and $C(\infty) = 0$.

The decay time of the ACF reflects the time scale of relaxation of its underlying property.

Here it as important to note, that the correlation time τ in the definition of C is not identical to the simulation time t. Rather, it describes a moving time interval that can be shifted on the t-axis of the simulation trajectory. The calculation of $C(\tau)$ involves the average denoted by $\langle \dots \rangle$ over time origins $\tau = t$ of the simulation trajectory. Auto-correlation functions can be used to calculate transport coefficients from simulation data sampled in an equilibrium state through the Green-Kubo integrals.[66, 67] This is useful since equilibrium simulations are usually easier to set-up than non-equilibrium simulations.

An example of a Green-Kubo calculation is the calculation of the self-diffusion coefficient D from the velocity autocorrelation function:[68]

$$D = \frac{1}{3} \int_0^\infty \langle \mathbf{v}(0) \cdot \mathbf{v}(\tau) \rangle d\tau \tag{2.30}$$

2.4 Coarse-Graining in Molecular Simulation

Even with the computational resources available today, the time and length scales of atomistically resolved molecular simulations are limited to scales of a few microseconds in simulation time and a few hundred nanometers in dimension. For simulations that require larger systems or longer simulation times, the use of models with lower resolution, a coarse-grained (CG) model, can be used.[7–10, 69] Lowering the model's resolution reduces the number of interaction sites required to model the same system and this makes a simulation with this model more efficient with regard to the required computational effort. In addition to the reduction of degrees of freedom in the system, CG models show intrinsically faster relaxation time scales than atomistic models[35, 69] and this adds to the increased computational performance achieved by CG models. The CG model will, by its very nature, lack some chemical detail in comparison with finer representations of the same system. Hence, the ideal degree of coarse-graining is determined by the nature of the problem at hand and the chemical accuracy the modeler needs to reproduce with the model.

With this limitation in mind, CG simulations can be used to model many different physical and physicochemical processes and phenomena on larger time and length scales than an atomistic model would allow. They can also be used to equilibrate large systems far from equilibrium into which atomistic detail may later be reinserted[29, 36], or serve as bath in adaptive resolution simulations, in which a smaller subregion is atomistically resolved.[70, 71]

Fine-grained (FG) and CG models of the same system can be related to each other using a mapping. The definition of a mapping is usually the first step in the development of a CG model. Formally the relation between a FG model with n_{FG} degrees of freedom (DoF) and a CG model with n_{CG} DoF is described by means of the mapping operator M. This operator acts on the n_{FG}-dimensional FG model coordinate vector $\mathbf{r}$, returning the n_{CG}-dimensional CG model coordinates $\mathbf{R}$:

$$\mathbf{R} = \mathbf{M}\mathbf{r} \tag{2.31}$$

In many cases the mapping is defined such that neighboring atoms are grouped together to form a CG interaction site. These atom groups are colloquially referred to as "blobs" or "beads". The choice of the mapping operator is very important for the behavior of the resulting CG model. In CG models based on atomistic models for soft matter the mapping is mostly performed as center of mass mapping of groups of related atoms.[10] The position of the interaction site α is then simply calculated by mass-weighing the positions of its constituent atoms i:

$$\mathbf{R}_\alpha = \sum_{i \in \alpha} \frac{m_i}{m_\alpha} \mathbf{r}_i, \tag{2.32}$$

where m_i is the individual atom mass and $m_\alpha = \sum m_i$ is the mass of the CG interaction site. Other mappings (e.g. center of geometry or center of charge) can also be used[24, 59], but they are far less common and are not considered in this thesis. In recent studies, several authors have discussed ideas to rationalize and optimize the choice of mapping by the use of modern machine learning methods.[72–75]

The transition from a fine-grained (FG) reference model to a CG model for the same system is referred to as coarse-graining. After the mapping scheme has been chosen, the interaction potentials of the CG model have to be formulated. Two paradigms exist for the parameterization of CG potentials: top-down and bottom-up.[69] Top-down models aim at the reproduction of macroscopic properties and achieve this by tuning the parameters of interaction potentials with a fixed functional form until the target properties are predicted by the CG model in agreement with a FG or experimental reference.[58, 76, 77] The focus of this thesis is on bottom-up methods which aim to reproduce the physics of the FG model by determining suitable CG interaction potentials through systematic evaluation of FG simulations.[8] Some of the commonly used bottom-up coarse-graining methods in the literature as well as the methods used in this thesis will be introduced briefly after a general discussion on the quality of CG models.

2.4.1 Representability and Transferability

The free energy of a CG system as a function of the configuration $\mathbf{R}$, the potential of mean force U_{PMF} is clearly defined through the underlying fine-grained interaction potential:[8]

$$U_{\mathrm{PMF}}(\mathbf{R}) = -\beta^{-1} \ln\left(\int \exp(-\beta U_{\mathrm{FG}}(\mathbf{r})) \delta(\mathbf{R} - \mathbf{M}\mathbf{r}) d\mathbf{r} \right) + C, \tag{2.33}$$

where $\beta = (k_{\mathrm{B}}T)^{-1}$ with Boltzmann constant k_{B} and temperature T, and C is a constant. The definition above defines the "complete" CG interaction free energy as an average over the potential energy of the FG micro-states $\mathbf{r}$ that can be associated with the CG macro-state $\mathbf{R}$ using the Dirac delta function δ to select the compatible micro-states. This "ideal" CG interaction free energy will be referred to as multi-body potential of mean force (multi-body PMF) from now on. Its definition is exact, but unfortunately it cannot be used as interaction potential in a CG simulation because, as a consequence of its multi-body character, it can not be computed but for systems with a few degrees of freedom, let alone be used as the basis for simulation algorithm.[10, 13] Consequently, CG models rely on an attempt to approximate the multi-body free energy surface by means of potentials which are of lower order.[8, 13, 78, 79] The functional forms chosen are usually similar to those chosen in the fine-grained simulations: nonbonded

interactions are modeled using pair interactions and bonded interactions of higher order are added where necessary to retain molecular geometry. The approximation of multi-order potentials through effective pair interactions is challenging and the properties of the CG model heavily depend on the choices made for this approximation.[8, 10]

One measure by which the quality of a CG model can be judged is the representability. A model is considered representable if the properties predicted by it are in accordance with those predicted by the FG reference model at the thermodynamic state point at which the CG model was parameterized.[11] A model is called representable if it uses a good approximation of the multi-body PMF as described above.

A major challenge of the simulation with coarse-grained models lies in parametrizing models that reproduce correct physical behavior not only at the state point of parameterization but also at different state points. This quality of a CG model is called transferability. Transferable CG models are of course very desirable, since CG model paramterization can be computationally expensive and in many applications some state variables (e.g. pressure or temperature) will change during a single simulation. Transferability and representability are linked to each other as there are properties of a system which require both qualities to be reproduced correctly.[23] It has been linked to the correct description of the entropy-enthalpy balance in the approximated potential of mean force.[13, 79]

2.4.2 Bottom-Up Coarse-Graining Approaches

Structure-Based Methods

Structure-based methods make use of structural relations in the systems to be modeled.[12, 78] For the parameterization of nonbonded interactions, the radial distribution function $g_{\mathrm{map}}(r)$ of the mapped FG system is commonly used. It describes a locally resolved relative density in an infinitesimal spherical shell with radius r around a CG site. Since it is essentially a distribution function with respect to the pair distance it can be Boltzmann inverted to calculate a corresponding free energy:

$$U_0(r) = -k_{\mathrm{B}}T \ln(g_{\mathrm{map}}(r)) \tag{2.34}$$

The resulting energy U_0 is a pairwise approximation to the PMF, albeit a rather crude one, since terms of higher-order contribute to the $g(r)$. This means that a CG simulation using U_0 will not reproduce the fine-grained reference structure $g_{\mathrm{map}}(r)$ but will produce a different structure $g_0(r)$.[12, 78]

In structure-based coarse-graining methods a correction term ΔU is calculated based on the mismatch between the target structure and the structure predicted by the CG interaction potential. This correction term is used to generate an updated potential U_1, which can is in turn used to perform a simulation and calculate a RDF $g_1(r)$. This procedure is iterated until the difference between the reference structure and the structure predicted by the CG model is below a specified threshold.

$$\Delta U_n = f(g_n(r), g_{\mathrm{map}}(r)) \tag{2.35}$$

$$U_{n+1} = U_n + \Delta U_n \tag{2.36}$$

The nature of the function used to relate changes in the potential and mismatches in the structure depends on the method used. In iterative Boltzmann inversion (IBI) it reads:[12]

$$\Delta U_n = k_{\mathrm{B}}T \ln\left(\frac{g_n(r)}{g_{\mathrm{map}}(r)}\right), \tag{2.37}$$

while in the inverse Monte-Carlo method (IMC) an exact Newton formulation is used:[80]

$$\Delta U_n = A^{-1}(g_n(r) - g_{\mathrm{map}}(r)), \tag{2.38}$$

where is the Jacobi matrix $A = \frac{\partial g}{\partial U}$.

In order to increase the transferability and representability of the CG potentials parameterized by the use of inverse methods, additional target properties can be used or information from different state points may be included.[16, 18, 81, 82]

Force-Matching

The force-matching approach to coarse-graining, also referred to as the multi-scale coarse-graining (MS-CG) method, also aims at the reproduction of a mapped property calculated from FG reference simulations.[10, 79, 83] Here the sum of the atomistic forces on the CG interaction site I $\mathbf{f}_I$ are compared to the forces on the same site predicted by the CG potential function $\mathbf{F}_I$. The parameterization of the CG potential function is not undertaken by iterating CG simulations but through minimization of the residual χ^2:

$$\chi^2 = \frac{1}{3N}\langle|\mathbf{F}_I(\mathbf{R}) - \mathbf{f}_I(\mathbf{r})|^2\rangle \tag{2.39}$$

The equation above can be solved for $\mathbf{F}_I(\mathbf{R})$ using a linear least squares optimization in cases were a suitable functional form of the CG potential (e.g. spline functions) can be asserted.[79] This makes the method non-iterative, which gives it a benefit in computational performance over the iterative inverse methods introduced before. A sufficiently large FG trajectory is sufficient to extract the CG interaction potential in one post-processing computation.

The MS-CG method has a solid grounding in statistical mechanics and it has been shown that the pair potentials derived with it yield a very good approximation of the multi-body PMF at the state point of parameterization.[79] Additional thermodynamic or structural properties can be included into the score function χ^2 to enhance the representability and transferability.[84]

A recent method based on the force-matching paradigm is the Ultra-Coarse-Graining (UCG) method [23] in which transferability can be successfully included into force-matching models by adding information on relevant substates of the FG model into the description of the CG interaction sites.[85–87]

Pair-Interaction Based Modeling

The coarse-graining methods presented so far share in common, that they attempt an approximation of the many-body PMF by pair forces and in doing so project atomistic interactions of many-body nature onto CG pair interactions.[7, 8, 10] In principle, this leads to a model which can accurately represent static properties at the state point of parameterization but is not very transferable with respect to changes in the system's state point.[24, 25] Many attempts have been made to make these models more transferable by adding optimization targets or sampling states to the parameterization procedure.[12–23]

Another methodological approach to achieving transferability is to generate a CG pair interaction by averaging only those atomistic interactions between atoms i and j that are part of the respective part of the CG site pair a and b. Formally, an atomistic force $\mathbf{F}_{ij}$ will contribute to the CG force $\mathbf{F}_{ab}$ only if $i \in a$ and $j \in b$. Interactions across the boundaries of CG interaction sites, $\mathbf{F}_{kj}$ and $\mathbf{F}_{il}$ with $k \notin a$ and $l \notin b$, are not included into $\mathbf{F}_{ab}$ since it is assumed that these interactions are the cause of multi-body dependencies between the CG interactions and consequently lead to a worse transferability of the resulting CG model.[24, 25]

The two methods employing this procedure, the effective-force coarse-graining (EFCG)[25] and conditional reversible work (CRW)[24] methods, produce CG models which are transferable with respect to temperature and the change in environment upon transferring a molecule from the bulk liquid to the vapor-liquid interface.[8] In these models the degree of coarse-graining is rather low and the interaction CG interaction potentials are still similar to those of the underlying FG model, the CG model could therefore also be described as a "united-atom model". Still, the coarse-graining of 2-6 heavy atoms into one CG site can already produce a significant increase in computational performance. In addition, the pair-interaction-based coarse-graining methods are computationally inexpensive and can be used to

vary mapping schemes and learn about the essential features of the FG model that need to be retained in order to obtain a representable and transferable CG model.

In the following, the EFCG and CRW methods are presented briefly and it is demonstrated that they are almost equal from the methodological point of view.

Effective-Force Coarse-Graining

The EFCG method has been developed by Wang et al. from the MS-CG coarse-graining framework.[25] A CG model is parameterized using a FG simulation of the system to be studied. Using the CG mapping, the norm of the CG interaction force $F_{ab}(r) = |\mathbf{F}_{ab}(r)|$ at constant distance r is calculated as a projection of the respective FG pair forces onto the distance vector between the CG sites:

$$F_{ab}(r) = \left\langle \sum_{i\in a} \sum_{j\in b} \mathbf{F}_{ij} \cdot \frac{\mathbf{R}_{ab}}{r} \right\rangle_r , \tag{2.40}$$

where $\mathbf{R}_{ab} = \mathbf{R}_a - \mathbf{R}_b$ is the distance vector between the CG sites of interest. The angular brackets $\langle \ldots \rangle_r$ denote an ensemble average over all CG site pairs at distance $r = |\mathbf{R}_{ab}|$.

The CG potential is calculated using these mean effective forces between the CG sites:

$$U_{\mathrm{EF-CG}}(r) = -\int_{\infty}^{r} F_{ab}(r')dr' \tag{2.41}$$

The EF-CG method can be formulated as a mean-field approximation to the MS-CG method.[25] The initial motivation for this method was that for small degrees of coarse-graining, EF-CG potentials, while not as accurate in representing the thermodynamics of the parameterization state, are more transferable with respect to the environment and temperature.[25, 26]

Whenever a FG trajectory is accessible the EF-CG potentials can be straightforwardly computed from reprocessing the FG position trajectory. It has been applied to study ionic liquids and successfully reproduces to a good degree the bulk density of the liquid, the thermal expansion coefficient, the vapor-liquid surface tension, and the temperature transferability of the $g(r)$.[26]

Conditional Reversible Work

Another method which is based on averaged pair interactions from FG simulations is the CRW method.[24] It has been developed initially to parametrize a CG model for polystyrene.[27] The equilibration of atomistically modeled polymer melt systems is computationally expensive making it difficult to sample a long enough trajectory to use the methods above which all necessitate the existence of an FG trajectory of the system to be studied. As a solution to this problem the CG interactions were sampled by simulations of molecule pairs in vacuo. The CRW method has afterwards also been applied with success for the parameterization of molecular liquids[24, 28, 30], penetrants in polymer melts[29, 36], and interactions between liquids and a graphene surface.[88]

The basis of a model parameterization with the CRW method are simulations of pairs of molecules *in vacuo*. A pair of molecules is placed such that the CG sites of interest are at a distance r_{C}. Throughout the length of the simulation this distance is kept constant, such that the constraint $|\mathbf{R}_{ab}| = r_{\mathrm{C}}$ is always satisfied. The projection of the forces between the molecules onto the CG site distance vector, the constraint force F_{C}, is calculated by summing all the underlying atomistic pair forces.

$$F_{\mathrm{C}}(r_{\mathrm{C}}) = \left\langle \sum_{i\in \mathrm{Mol.1}} \sum_{j\in \mathrm{Mol.2}} \mathbf{F}_{ij} \cdot \frac{\mathbf{R}_{ab}}{r_{\mathrm{C}}} \right\rangle, \tag{2.42}$$

Integration of F_{C} yields the potential energy equivalent to the work required to reversibly change the distance between the pair of molecules:

$$W(r) = -\int_{\infty}^{r_{\mathrm{C}}} F_{\mathrm{C}} dr_{\mathrm{C}} \tag{2.43}$$

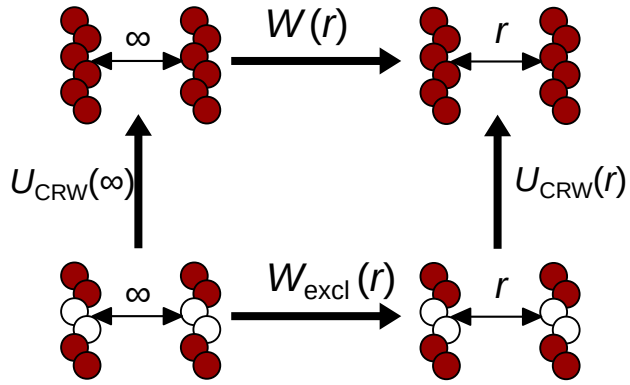


Figure 2.1: The approach of the CRW method can be illustrated using a thermodynamic cycle. The aim is to find the CG interaction free energy $U_{\mathrm{CRW}}(r)$. This quantity can be determined from the reversible work required to displace two molecules from quasi-infinite distance to the distance r. This work is calculated twice, first with the full set of atomistic interactions enabled ($W(r)$) and second with all interactions between atoms contained in the CG sites of interest excluded ($W_{\mathrm{excl}}(r)$).

This quantity can be called a potential of mean force (PMF) but it should not be confused with the rigorous multi-body PMF which describes the ideal CG free energy surface.

In order to extract an interaction potential between the CG sites a and b, the simulations at constraint distance are repeated using a different set of interactions in the FG model. In the second simulation all pair interactions between atoms $i \in a$ and $j \in b$ are excluded. These simulations will yield a second PMF W_{excl}.

Assuming that the interaction between CG sites is zero at quasi-infinite distance, the CG interaction potential can be calculated as difference of the two PMFs:

$$U_{\mathrm{CRW}}(r) = W(r) - W_{\mathrm{excl}}(r). \tag{2.44}$$

The CRW interaction potential can also be interpreted as the free energy change of the process of reversibly "switching-on" the pair interactions between the two CG sites. This process is instructively illustrated using a thermodynamic cycle in figure 2.1.

The resulting CG potential is called conditional because it is parameterized *under the condition* that the CG site is embedded in its chemical surroundings.[24] It has been pointed out that especially for large molecules the vicinity of neighboring CG sites leads to a good estimation of the entropy part of the interaction free energy.[8, 28] This entropical contribution is a measure of the change in configurational volume accessible to the CG sites upon inserting the direct interactions between the pair of CG sites of interest.[24] A correct representation of the entropical contribution to the interaction free energy has been linked to good transferability of the CG model. Indeed, CRW models have been shown to be very transferable at the expense of a less accurate representation of the parameterization state point.[24, 27–30]

Similarity of EF-CG and CRW Potentials

It is shown in the following that the potentials calculated with the EF-CG and CRW methods are very similar in principle. Nonwithstanding the methodical similarity, major differences in potentials generated from the two methods will in most applications arise from the fact that CRW is most efficiently performed in vacuo while EF-CG is more often performed using a trajectory sampled in the liquid state. In the following it will be assumed that the sampling state is identical for both methods.

Recall the definition of the EF-CG interaction potential:

$$U_{\mathrm{EF-CG}}(r) = -\int_{\infty}^{r} F_{ij}(r')dr', \tag{2.45}$$

where $F_{ij} = \langle \sum_a \sum_b F_{ab} \rangle$ is the projection onto the i-j unit vector of the sum of the atomistic pair forces between atoms a and b contained in sites i and j, respectively. Adding and subtracting to F_{ij} the 'complement' force F_{excl} (already known from the CRW method):

$$F_{\text{excl}} = \left\langle \sum_{x \in \text{Mol.1}} \sum_{y \in \text{Mol.2}} F_{xy} \xi_{xiyj} \right\rangle \tag{2.46}$$

$$\text{with } \xi_{xiyj} = \begin{cases} 0 \text{ if } (x \in i) \wedge (y \in j) \\ 1 \text{ if } (x \notin i) \vee (y \notin j) \end{cases}, \tag{2.47}$$

yields

$$F_{ij} = F_{ij} + F_{\text{excl}} - F_{\text{excl}}, \tag{2.48}$$

which can be substituted into eq. 2.45:

$$-U_{\text{EF-CG}}(r) = \int_{\infty}^{r} F_{ij} dr' \tag{2.49}$$

$$= \int_{\infty}^{r} F_{ij}(r') + F_{\text{excl}}(r') dr' - \int_{\infty}^{r} F_{\text{excl}}(r') dr' \tag{2.50}$$

$$= -W(r) - \int_{\infty}^{r} F_{\text{excl}}(r') dr', \tag{2.51}$$

where the integral over the sum of F_{ij} and F_{excl} has been identified as the potential of mean force used in the definition of the CRW potential (eq. 2.44).

Further, the second integral term can be expressed as an approximation to W_{excl} used in eq. 2.44:

$$\int_{\infty}^{r} F_{\text{excl}}(r') dr' \approx -W_{\text{excl}}. \tag{2.52}$$

The relation between these two terms is not an equality because the PMF W_{excl} is calculated from a trajectory sampled with the direct interactions between atoms in i and j excluded. The mean force F_{excl} on the other side is sampled from a trajectory with all forces included. It is not straightforward to estimate the differences in CRW and EF-CG potentials that arise from this difference in sampling, but if they can be assumed to be negligeable (which should be the case especially for low degrees of coarse-graining[28]) we can conclude that the EF-CG and CRW potentials are approximately equal, if the sampling environment used in their parameterization is identical:

$$U_{\text{EF-CG}}(r) \approx W(r) - W_{\text{excl}}(r) = U_{\text{CRW}} \tag{2.53}$$

2.5 Coarse-Graining and Dynamics

CG models are successfully used for the study of thermodynamic and structural properties. These quantities are ensemble averages which only depend on the Hamiltonian used to describe the system. Molecular dynamics also allows the study of dynamic quantities through observation of the time evolution of molecular positions and momenta. However, CG models usually fail in the reproduction of these quantities. The characteristic time scale of dynamic modes in a CG simulation is typically faster than the time scale in an atomistic simulation.[35]

A CG model which aims at the consistent representation of FG time scales has to correctly capture the effects of the coarse-graining procedure on the time evolution of the phase space variables. This requires

not only the molecules and interaction potentials to be coarse-grained but also requires a coarse-grained equation of motion (EOM), which can be systematically derived using the Mori-Zwanzig Projection operator formalism.

In this section the time evolution in the CG model is analyzed and the derivation of a CG EOM is presented that adds to the conservative term from the CG interaction potential, dissipative and noise contributions which allow for a correction of the "speed-up" of internal dynamics caused by the flattening of the potential energy surface in the CG model.

2.5.1 The Liouville Operator and the Propagation Operator

A system of atoms with phase space coordinates $\Gamma = \{\mathbf{r}, \mathbf{p}\}$ (including the position and momentum vectors $\mathbf{r}$ and $\mathbf{p}$ of length $3N$) can be described using the Hamiltonian $\mathscr{H}$:

$$\mathscr{H} = \sum_{j}^{3N} \frac{p_j^2}{2m_j} + U(\mathbf{r}) \tag{2.54}$$

In this system the time evolution of a variable $X(\mathbf{r}, \mathbf{p})$ that depends on the phase space coordinates can be expressed by mean of the Liouville operator $i\mathscr{L}$:[68]

$$\frac{dX}{dt} = i\mathscr{L}X(\mathbf{r}, \mathbf{p}) = i\sum_{j}^{N_j} \left(\frac{\partial \mathscr{H}}{\partial \mathbf{r}_j} \frac{\partial}{\partial \mathbf{p}_j} + \frac{\partial \mathscr{H}}{\partial \mathbf{p}_j} \frac{\partial}{\partial \mathbf{r}_j} \right) X \tag{2.55}$$

Using the Liouville operator, the time evolution of X can be described using a Taylor expansion around the time origin $t = 0$:[68]

$$X(t) = \exp(i\mathscr{L}t)X(0) \tag{2.56}$$

The operator $\exp(i\mathscr{L}t)$ is called propagation operator or propagator.

In coarse-grained systems, a mapping scheme is employed to reduce the number of degrees of freedom, i.e. the number of relevant positions and momenta. Such a system can be described by decomposing the phase space coordinate Γ of the underlying fine-grained system such that we have N_{CG} relevant and N_{relative} irrelevant degrees of freedom. If a center-of-mass based mapping is performed and the irrelevant degrees of freedom are chosen in a canonical manner ($N = N_{\mathrm{CG}} + N_{\mathrm{relative}}$), the Liouville operator can be decomposed into a slow (describing the CG DoF) and a fast (describing the relative DoF) component ($\mathscr{L}_s$ and $\mathscr{L}_f$, respectively):[40]

$$\mathscr{L} = \mathscr{L}_s + \mathscr{L}_f = \sum_{i}^{N_{CG}} \left(\mathbf{F}_i \frac{\partial}{\partial \mathbf{P}_i} + \mathbf{P}_i \frac{\partial}{\partial \mathbf{R}_i} \right) + \sum_{i}^{N_r} \left(\frac{\partial \mathscr{H}}{\partial \mathbf{r}_i^r} \frac{\partial}{\partial \mathbf{p}_i^r} + \frac{\partial \mathscr{H}}{\partial \mathbf{p}_i^r} \frac{\partial}{\partial \mathbf{r}_i^r} \right) \tag{2.57}$$

The propagator can also be described using these two components. On a basic level this explains the speed-up of dynamics in (conservative) coarse-grained models. Here only the slow component of the Liouville operator is used to propagate the system. The aim of a dynamically consistent CG model is to perform the propagation through phase space while taking into account also the effect of the fast DoF on the slow DoF, without explicitly modeling all fine-grained sites. In the following it will be shown that an equation of motion for such a system can be obtained through the so-called projection operator formalism. This equation of motion is a generalized Langevin equation (GLE) that includes a dissipative and a random term in addition to the conservative force used in simple CG models.

2.5.2 The Projection Operator Formalism

Since its introduction by Nakajima[37], Mori[39] and Zwanzig[38], the projection operator formalism has been used to describe the time evolution of systems with reduced complexity. The underlying idea is the following: a projection operator (PO) $\mathscr{P}$ is defined that projects a vector onto a subspace of the total vector space. In addition, the complementary operator $\mathscr{Q} = \mathscr{I} - \mathscr{P}$, where $\mathscr{I}$ is the identity operator, is used to calculate the remaining component of the vector. The formal definition of the projection operator to project a quantity X onto a direction A reads as follows:

$$\mathscr{P}X = \frac{(A,X)}{(A,A)}A. \tag{2.58}$$

Here $(.,.)$ denotes the inner product which can be replaced by the ensemble weighted average:[89]

$$(A(\mathbf{\Gamma}),B(\mathbf{\Gamma})) = \int d\mathbf{\Gamma}\rho(\mathbf{\Gamma})A(\mathbf{\Gamma})B(\mathbf{\Gamma}), \tag{2.59}$$

where ρ describes the phase space density of the state point denoted Γ.

Before applying the PO formalism to a general system, the properties of the PO are illustrated using simple vectors in the 3D space.

Projections in 3D vector space

Imagine arbitrary 3D vectors $\mathbf{x}$ and $\mathbf{r}$. In geometric terms $\mathbf{x}$ can be decomposed into two orthogonal components $\mathbf{x}_{\parallel r}$ and $\mathbf{x}_{\perp r}$ that are parallel and perpendicular to $\mathbf{r}$ respectively. These parallel component is calculated as follows:

$$\mathbf{x}_{\parallel r} = (\mathbf{e}_r \cdot \mathbf{x})\mathbf{e}_r, \tag{2.60}$$

using a unit vector $\mathbf{e}_r$ parallel to $\mathbf{r}$. Clearly $\mathbf{e}_r = \mathbf{r}/|r|$ and $|r|^2 = \mathbf{r}\cdot\mathbf{r}$, so the calculation of the parallel component can be rewritten as follows:

$$\mathbf{x}_{\parallel r} = \frac{\mathbf{r}\cdot\mathbf{x}}{\mathbf{r}\cdot\mathbf{r}}\mathbf{r} = \mathscr{P}\mathbf{x} \tag{2.61}$$

From this relation the dot product can be identified as the inner product from the definition of the PO for the 3D vectors and it has been shown that the PO returns the parallel component along $\mathbf{r}$. The complement $\mathbf{x}_{\perp r}$ is calculated through:

$$\mathbf{x}_{\perp r} = \mathbf{x} - (\mathbf{e}_r\cdot\mathbf{x})\mathbf{e}_r = \mathbf{x} - \mathscr{P}\mathbf{x} = (\mathscr{I}-\mathscr{P})\mathbf{x} = \mathscr{Q}\mathbf{x}, \tag{2.62}$$

using the identity operator $\mathscr{I}$.

The projection operators have a few interesting properties, which will be shown in the following. Firstly, the POs are idempotent, that is $\mathscr{P} = \mathscr{P}\mathscr{P}$ which is easily shown:

$$\mathscr{P}\mathscr{P}\mathbf{x} = \mathscr{P}\mathbf{x}_{\parallel r} = \mathbf{x}_{\parallel r} \tag{2.63}$$

where the second projection does not change the result, because $\mathbf{x}_{\parallel r}$ is already parallel to $\mathbf{r}$.

A second interesting property of the projection operator is its self-adjointness with respect to the inner product:

$$(\mathscr{P}\mathbf{x})\cdot\mathbf{y} = \mathbf{x}\cdot(\mathscr{P}\mathbf{y}), \tag{2.64}$$

where $\mathbf{y}$ is a second arbitrary vector. The proof of this property is straightforward:

$$(\mathscr{P}\mathbf{x})\cdot\mathbf{y} = \frac{1}{r^2}((\mathbf{r}\cdot\mathbf{x})\mathbf{r})\cdot\mathbf{y} = \frac{1}{r^2}(\mathbf{r}\cdot\mathbf{x})(\mathbf{r}\cdot\mathbf{y}) = \mathbf{x}\cdot\frac{1}{r^2}(\mathbf{r}\cdot\mathbf{y})\mathbf{r} = \mathbf{x}\cdot(\mathscr{P}\mathbf{y}) \tag{2.65}$$

Of course, these two properties of $\mathscr{P}$ also apply $\mathscr{Q}$. Furthermore, the orthogonal nature of $\mathscr{P}$ and $\mathscr{Q}$ dictates, that:

$$\mathscr{P}\mathscr{Q}\mathbf{x} = \mathscr{Q}\mathscr{P}\mathbf{x} = 0 \tag{2.66}$$

These properties apply also in the more general case used in the following derivation of the CG equation of motion.

Projection operators acting on phase space observables

Imagine a system with phase space coordinates:

$$\Gamma = \{\Gamma_{\mathrm{CG}}, \mathbf{\Gamma}_{\mathrm{r}}\} = \{\{\mathbf{R},\mathbf{P}\},\{\mathbf{r}_{\mathrm{r}},\mathbf{p}_{\mathrm{r}}\}\}, \tag{2.67}$$

with a number of relevant (CG) degrees of freedom and a number of irrelevant (relative,r) DoF. Imagine further a phase variable $X(\Gamma(t))$ that depends on time not explicitly but only through the phase space coordinates of the system. In the coarse-graining application this variable is obviously the CG phase space coordinate Γ_{CG} which depends on all FG phase space variables. The aim of the PO formalism in coarse-graining applications is the extraction of a formally exact EOM for these CG coordinates based on an observation of the time evolution of the underlying FG coordinates. The time evolution of Γ_{CG} is described by the Liouville operator $i\mathscr{L}$ of the atomistic system. The slow contribution of the time evolution can be obtained by projecting the Liouvillian onto the set of relevant variables and the fast contribution can be obtained through the orthogonal complement thereof:

$$\frac{d\Gamma_{\mathrm{CG}}}{dt} = i\mathscr{L}\Gamma_{\mathrm{CG}} = i(\mathscr{P}+\mathscr{Q})\mathscr{L}\Gamma_{\mathrm{CG}} \tag{2.68}$$

The slow part of the dynamics $i\mathscr{P}\mathscr{L}\Gamma_{\mathrm{CG}}$ describes the conservative part of the CG equation of motion. Using only this part of the EOM for a simulation will lead to the speed-up that is often observed in CG systems.[33]

In a dynamically consistent CG equation of motion the effect of the fast part on the CG dynamics also has to be taken into account. The time derivative of Γ_{CG} is (like Γ_{CG} itself) time-dependent through the phase variables $\mathbf{\Gamma}$. Hence, it can be propagated using the propagator defined in eq. 2.56:

$$\frac{d\Gamma_{\mathrm{CG}}}{dt}(t) = \exp(i\mathscr{L}t)i\mathscr{L}\Gamma_{\mathrm{CG}}(0) = \exp(i\mathscr{L}t)\left(i\mathscr{P}\mathscr{L}\Gamma_{\mathrm{CG}}(0) + i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}(0)\right) \tag{2.69}$$

The second term can be rewritten through the Duhamel-Dyson identity [34, 41, 68]:

$$\begin{aligned}&\exp(i\mathscr{L}t)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}(0)\\ &= \int_0^t d\tau\,\exp(i\mathscr{L}(t-\tau))i\mathscr{P}\mathscr{L}\exp(i\mathscr{Q}\mathscr{L}\tau)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}(0) + \exp(i\mathscr{Q}\mathscr{L}t)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}(0)\end{aligned} \tag{2.70}$$

The second term herein is obtained solely through the Q-projector and it is therefore always orthogonal to the slow DoF. It has been identified therefore as noise that does not influence the average properties of the system, but acts as an oscillating random term. For this reason it will be called random force ($\delta\mathbf{F}^Q(t)$) from now on:

$$\delta\mathbf{F}^Q(t) \equiv \exp(i\mathscr{Q}\mathscr{L}t)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}} = \mathscr{Q}\exp(i\mathscr{Q}\mathscr{L}t)i\mathscr{L}\Gamma_{\mathrm{CG}} \tag{2.71}$$

The choice of the description "random *force*" for this fluctuating term will become more clear in a moment.

The meaning of the integral term in eq. 2.70 is more complicated as it contains both projections using $\mathscr{P}$ and $\mathscr{Q}$. As a first simplification, the definition of $\delta\mathbf{F}^Q(t)$ from eq. 2.71 can be substituted into the expression to yield:

$$\exp(i\mathscr{L}t)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}(0)-\delta\mathbf{F}^Q(t)=\int_0^t d\tau\exp(i\mathscr{L}(t-\tau))i\mathscr{P}\mathscr{L}\delta\mathbf{F}^Q(\tau) \tag{2.72}$$

The following, further simplifications make use of the operator properties described earlier:

- Idempotence of the projection operators: $\mathscr{Q}\delta\mathbf{F}^Q=\delta\mathbf{F}^Q$
- Anti-self-adjointness of the Liouvillean:[68] $\langle i\mathscr{L}AB\rangle=-\langle Ai\mathscr{L}B\rangle$
- Self-adjointness of the projection operators: $\langle\mathscr{Q}AB\rangle=\langle A\mathscr{Q}B\rangle$

and will lead to a better understanding of the meaning of the integral term of equation 2.70:

$$\exp(i\mathscr{L}t)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}(0)-\delta\mathbf{F}^Q(t)=\int_0^t d\tau\exp(i\mathscr{L}(t-\tau))\frac{\langle i\mathscr{L}\mathscr{Q}\delta\mathbf{F}^Q(\tau)\Gamma_{\mathrm{CG}}\rangle}{\langle{\Gamma_{\mathrm{CG}}}^2\rangle}\Gamma_{\mathrm{CG}} \tag{2.73}$$

$$=-\int_0^t d\tau\exp(i\mathscr{L}(t-\tau))\frac{\langle\mathscr{Q}\delta\mathbf{F}^Q(\tau)i\mathscr{L}\Gamma_{\mathrm{CG}}\rangle}{\langle{\Gamma_{\mathrm{CG}}}^2\rangle}\Gamma_{\mathrm{CG}} \tag{2.74}$$

$$=-\int_0^t d\tau\exp(i\mathscr{L}(t-\tau))\frac{\langle\delta\mathbf{F}^Q(\tau)i\mathscr{Q}\mathscr{L}\Gamma_{\mathrm{CG}}\rangle}{\langle{\Gamma_{\mathrm{CG}}}^2\rangle}\Gamma_{\mathrm{CG}} \tag{2.75}$$

$$=-\int_0^t d\tau\exp(i\mathscr{L}(t-\tau))\frac{\langle\delta\mathbf{F}^Q(\tau)\delta\mathbf{F}^Q(0)\rangle}{\langle{\Gamma_{\mathrm{CG}}}^2\rangle}\Gamma_{\mathrm{CG}} \tag{2.76}$$

$$=-\int_0^t d\tau\frac{\langle\delta\mathbf{F}^Q(\tau)\delta\mathbf{F}^Q(0)\rangle}{\langle{\Gamma_{\mathrm{CG}}}^2\rangle}\Gamma_{\mathrm{CG}}(t-\tau) \tag{2.77}$$

It has been shown here that the integral term in the evolution of the projected dynamics is proportional to a time correlation function of the random force $\delta\mathbf{F}^Q$. This is the most important finding of the PO formalism and will enable the derivation of a practically useful EOM for the CG system.

Joining the terms derived above we find the MZ equation for the CG phase space variables:

$$\frac{d\Gamma_{\mathrm{CG}}}{dt}(t)=i\mathscr{P}\mathscr{L}\Gamma_{\mathrm{CG}}-\int_0^t d\tau\frac{\langle\delta\mathbf{F}^Q(\tau)\delta\mathbf{F}^Q(0)\rangle}{\langle{\Gamma_{\mathrm{CG}}}^2\rangle}\Gamma_{\mathrm{CG}}(t-\tau)+\delta\mathbf{F}^Q(t) \tag{2.78}$$

A coarse-grained equation of motion

For a use in a CG simulation the MZ equation 2.78 has to be analyzed to find terms for the time evolution of the CG positions $\mathbf{R}_a$ and momenta $\mathbf{P}_a$ of the CG sites a. The Liouvillean applied to the CG position vector reads:

$$\mathscr{L}\mathbf{R}_a=\frac{\partial\mathscr{H}}{\partial\mathbf{P}_a}=\frac{\mathbf{P}_a}{m_a} \tag{2.79}$$

The CG momentum vector is, by definition, a variable in the $\mathscr{P}$ subspace. Therefore the P-projection of the momentum vector is the momentum vector itself and the Q-projection is zero. As a consequence, the time evolution of the CG positions is described only through the P-projection:

$$\frac{d\mathbf{R}_a}{dt}=\frac{\mathbf{P}_a}{m_a} \tag{2.80}$$

The Q-projected part of the dynamics only exerts an influence in the time evolution of the CG momenta, i.e. the forces on the CG sites:

$$\frac{d\mathbf{P}_a}{dt}(t) = \mathbf{F}_a(t) = \mathscr{P}\mathscr{L}\mathbf{P}_a(t) - \int_0^t d\tau \frac{\langle \delta\mathbf{F}_a^Q(\tau)\delta\mathbf{F}_a^Q(0)\rangle}{\langle \mathbf{P}_a^2(0)\rangle}\mathbf{P}_a(t-\tau) + \delta\mathbf{F}_a^Q(t) \tag{2.81}$$

$$= \langle \mathbf{F}_a \rangle (t) - \frac{1}{k_B T m_a}\int_0^t d\tau \langle \delta\mathbf{F}_a^Q(\tau)\delta\mathbf{F}_a^Q(0)\rangle \mathbf{P}_a(t-\tau) + \delta\mathbf{F}_a^Q(t), \tag{2.82}$$

The latter is the generalized Langevin equation (GLE), where in the mean force $\langle \mathbf{F}_a \rangle$ is substituted for the P-projection of the force. It has been shown that this mean force is equivalent to the force derived from the multi-body potential of mean force $-\partial U_{\text{eff}}/\partial \mathbf{R}_a$, a property which has already been introduced earlier in the section on coarse-graining methodology (2.4.1).[34, 41] It can be concluded therefore that a CG model in which solely the (ideal multi-body) mean force is used as interaction force yields an inaccurate description of the time evolution in phase space, although the correct average values are predicted by it. For a model to describe dynamic properties correctly additional terms have to be included. These terms depend on the fluctuating forces $\delta\mathbf{F}_a^Q$ around the mean force.

The term random force, which was introduced earlier can now be justified since the Q-projected parts of the dynamics indeed only effect the forces in the CG model, not the velocities. The middle term in the GLE includes a time integral over past state points and it can hence be described as a memory term.

The GLE itself is however not very well suited to be used directly as a basis of molecular simulations. Further simplifications and assumptions can be made to enable CG simulations with a consistent dynamical behavior. This procedure is described in more detail in section 6.3.

2.6 References

1. Metropolis, N. & Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association* **44,** 335–341 (1949).
2. Alder, B. J. & Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics* **31,** 459–466 (1959).
3. Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications* 2nd ed. *Computational science series* **1** (Academic Press, San Diego, 2002).
4. Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids* Reprinted. OCLC: 845656692 (Clarendon Press, Oxford, 2009).
5. Perilla, J. R. *et al.* Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology* **31,** 64–74 (2015).
6. Marrink, S. J. *et al.* Computational Modeling of Realistic Cell Membranes. *Chemical Reviews* (2019).
7. Peter, C. & Kremer, K. Multiscale simulation of soft matter systems. *Faraday Discussions* **144.** 00053, 9 (2010).
8. Brini, E. *et al.* Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter* **9.** 00021, 2108 (2013).
9. Riniker, S., Allison, J. R. & van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Physical Chemistry Chemical Physics* **14,** 12423 (2012).
10. Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* **139,** 090901 (2013).
11. Dunn, N. J. H., Foley, T. T. & Noid, W. G. Van der Waals Perspective on Coarse-Graining: Progress toward Solving Representability and Transferability Problems. *Accounts of Chemical Research* **49,** 2832–2840 (2016).

12. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24.** 00428, 1624–1636 (2003).

13. Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129,** 144108 (2008).

14. Wang, H., Junghans, C. & Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *The European Physical Journal E* **28,** 221–229 (2009).

15. Mullinax, J. W. & Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *The Journal of Chemical Physics* **131,** 104110 (2009).

16. Farah, K., Fogarty, A. C., Böhm, M. C. & Müller-Plathe, F. Temperature dependence of coarse-grained potentials for liquid hexane. *Phys. Chem. Chem. Phys.* **13,** 2894–2902 (2011).

17. Ganguly, P., Mukherji, D., Junghans, C. & Van der Vegt, N. F. A. Kirkwood–Buff Coarse-Grained Force Fields for Aqueous Solutions. *Journal of Chemical Theory and Computation* **8.** 00013, 1802–1807 (2012).

18. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140,** 224104 (2014).

19. Rosenberger, D., Hanke, M. & Vegt, N. F. A. V. d. Comparison of iterative inverse coarse-graining methods. *The European Physical Journal Special Topics* **225,** 1323–1345 (2016).

20. Dunn, N. J. H. & Noid, W. G. Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures. *The Journal of Chemical Physics* **144,** 204124 (2016).

21. Wagner, J. W., Dannenhoffer-Lafage, T., Jin, J. & Voth, G. A. Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions. *The Journal of Chemical Physics* **147,** 044113 (2017).

22. Sanyal, T. & Shell, M. S. Transferable Coarse-Grained Models of Liquid–Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *The Journal of Physical Chemistry B* **122,** 5678–5693 (2018).

23. Jin, J. & Voth, G. A. Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *Journal of Chemical Theory and Computation* **14,** 2180–2197 (2018).

24. Brini, E., Marcon, V. & Van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics* **13.** 00022, 10468 (2011).

25. Wang, Y., Noid, W. G., Liu, P. & Voth, G. A. Effective force coarse-graining. *Physical Chemistry Chemical Physics* **11,** 2002 (2009).

26. Wang, Y., Feng, S. & Voth, G. A. Transferable Coarse-Grained Models for Ionic Liquids. *J. Chem. Theory Comput.* **5,** 1091–1098 (2009).

27. Fritz, D., Harmandaris, V. A., Kremer, K. & Van der Vegt, N. F. A. Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules* **42.** 00065, 7579–7588 (2009).

28. Brini, E. & Van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *The Journal of Chemical Physics* **137.** 00009, 154113 (2012).

29. Brini, E., Herbers, C. R., Deichmann, G. & Van der Vegt, N. F. A. Thermodynamic transferability of coarse-grained potentials for polymer–additive systems. *Physical Chemistry Chemical Physics* **14.** 00006, 11896 (2012).

30. Dallavalle, M. & Van der Vegt, N. F. A. Evaluation of mapping schemes for systematic coarse graining of higher alkanes. *Phys. Chem. Chem. Phys.* **19,** 23034–23042 (2017).

31. Tschöp, W., Kremer, K., Batoulis, J., Bürger, T. & Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica* **49.** 00310, 61–74 (1998).

32. Akkermans, R. L. C. & Briels, W. J. Coarse-grained dynamics of one chain in a polymer melt. *The Journal of Chemical Physics* **113,** 6409 (2000).

33. Izvekov, S. & Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *The Journal of Chemical Physics* **125,** 151101 (2006).

34. Hijón, C., Español, P., Vanden-Eijnden, E. & Delgado-Buscalioni, R. Mori–Zwanzig formalism as a practical computational tool. *Faraday Discussions* **144,** 301 (2010).

35. Fritz, D., Koschke, K., Harmandaris, V. A., Van der Vegt, N. F. A. & Kremer, K. Multiscale modeling of soft matter: scaling of dynamics. *Phys. Chem. Chem. Phys.* **13,** 10412 (2011).

36. Fritz, D., Herbers, C. R., Kremer, K. & Van der Vegt, N. F. A. Hierarchical modeling of polymer permeation. *Soft Matter* **5,** 4556 (2009).

37. Nakajima, S. On Quantum Theory of Transport Phenomena: Steady Diffusion. *Progress of Theoretical Physics* **20,** 948–959 (1958).

38. Zwanzig, R. Memory Effects in Irreversible Thermodynamics. *Physical Review* **124,** 983–992 (1961).

39. Mori, H. Transport, Collective Motion, and Brownian Motion. *Progress of Theoretical Physics* **33,** 423–455 (1965).

40. Izvekov, S. Microscopic derivation of particle-based coarse-grained dynamics. *The Journal of Chemical Physics* **138,** 134106 (2013).

41. Kinjo, T. & Hyodo, S. Equation of motion for coarse-grained simulation based on microscopic description. *Physical Review E* **75,** 051109 (2007).

42. Hoogerbrugge, P. J. & Koelman, J. M. V. A. Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics. *Europhysics Letters (EPL)* **19.** 01901, 155–160 (1992).

43. Español, P. & Warren, P. Statistical Mechanics of Dissipative Particle Dynamics. *Europhysics Letters (EPL)* **30.** 00000, 191–196 (1995).

44. Groot, R. D. & Warren, P. B. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *The Journal of Chemical Physics* **107.** 01852, 4423 (1997).

45. Español, P. & Warren, P. B. Perspective: Dissipative particle dynamics. *The Journal of Chemical Physics* **146,** 150901 (2017).

46. Eriksson, A., Jacobi, M. N., Nyström, J. & Tunstrøm, K. A method for estimating the interactions in dissipative particle dynamics from particle trajectories. *Journal of Physics: Condensed Matter* **21.** 00007, 095401 (2009).

47. Eriksson, A., Jacobi, M. N., Nyström, J. & Tunstrøm, K. Effective thermostat induced by coarse graining of simple point charge water. *The Journal of Chemical Physics* **129.** 00017, 024106 (2008).

48. Eriksson, A., Jacobi, M. N., Nyström, J. & Tunstrøm, K. Bottom-up derivation of an effective thermostat for united atoms simulations of water. *The Journal of Chemical Physics* **130.** 00008, 164509 (2009).

49. Lei, H., Caswell, B. & Karniadakis, G. E. Direct construction of mesoscopic models from microscopic simulations. *Physical Review E* **81.** 00039, 026704 (2010).

50. Li, Z., Bian, X., Caswell, B. & Karniadakis, G. E. Construction of dissipative particle dynamics models for complex fluids via the Mori–Zwanzig formulation. *Soft Matter* **10,** 8659–8672 (2014).

51. Izvekov, S. & Rice, B. M. Multi-scale coarse-graining of non-conservative interactions in molecular liquids. *The Journal of Chemical Physics* **140.** 00000, 104104 (2014).

52. Trément, S., Schnell, B., Petitjean, L., Couty, M. & Rousseau, B. Conservative and dissipative force field for simulation of coarse-grained alkane molecules: A bottom-up approach. *The Journal of Chemical Physics* **140,** 134113 (2014).

53. Deichmann, G., Marcon, V. & Van der Vegt, N. F. A. Bottom-up derivation of conservative and dissipative interactions for coarse-grained molecular liquids with the conditional reversible work method. *The Journal of Chemical Physics* **141,** 224109 (2014).

54. Li, Z., Bian, X., Li, X. & Karniadakis, G. E. Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism. *The Journal of Chemical Physics* **143,** 243128 (2015).

55. Dequidt, A. & Solano Canchaya, J. G. Bayesian parametrization of coarse-grain dissipative dynamics models. *The Journal of Chemical Physics* **143,** 084122 (2015).

56. Li, Z., Lee, H. S., Darve, E. & Karniadakis, G. E. Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts. *The Journal of Chemical Physics* **146,** 014104 (2017).

57. Tuckerman, M. E. *Statistical mechanics: theory and molecular simulation* OCLC: ocn551495372 (Oxford University Press, Oxford ; New York, 2010).

58. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).

59. Cao, Z. & Voth, G. A. The multiscale coarse-graining method. XI. Accurate interactions based on the centers of charge of coarse-grained sites. *J. Chem. Phys.* **143,** 243116 (2015).

60. Woodcock, L. Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters* **10,** 257–261 (1971).

61. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81.** 13416, 3684 (1984).

62. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50,** 1055–1076 (1983).

63. Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (1985).

64. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **17,** 1302–1322 (1978).

65. Junghans, C., Praprotnik, M. & Kremer, K. Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat. *Soft Matter* **4.** 00051, 156 (2008).

66. Green, M. S. Markoff Random Processes and the Statistical Mechanics of Time-Dependent Phenomena. II. Irreversible Processes in Fluids. *The Journal of Chemical Physics* **22,** 398–413 (1954).

67. Kubo, R. Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems. *Journal of the Physical Society of Japan* **12,** 570–586 (1957).

68. Evans, D. J. & Morriss, G. P. *Statistical mechanics of nonequilibrium liquids* 2. ed., 1. paperb. ed. OCLC: 894648614 (Cambridge Univ. Press, Cambridge, 2014).

69. Ingólfsson, H. I. *et al.* The power of coarse graining in biomolecular simulations: The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4,** 225–248 (2014).

70. Praprotnik, M., Delle Site, L. & Kremer, K. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *The Journal of Chemical Physics* **123,** 224106 (2005).

71. Nielsen, S. O., Bulo, R. E., Moore, P. B. & Ensing, B. Recent progress in adaptive multiscale molecular dynamics simulations of soft matter. *Physical Chemistry Chemical Physics* **12,** 12401 (2010).

72. Foley, T. T., Shell, M. S. & Noid, W. G. The impact of resolution upon entropy and information in coarse-grained models. *The Journal of Chemical Physics* **143,** 243104 (2015).

73. Bereau, T. & Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *Journal of Chemical Theory and Computation* **11,** 2783–2791 (2015).

74. Han, Y., Dama, J. F. & Voth, G. A. Mesoscopic coarse-grained representations of fluids rigorously derived from atomistic models. *The Journal of Chemical Physics* **149,** 044104 (2018).

75. Webb, M. A., Delannoy, J.-Y. & de Pablo, J. J. Graph-Based Approach to Systematic Molecular Coarse-Graining. *Journal of Chemical Theory and Computation* **15,** 1199–1208 (2019).

76. Binder, K. *et al.* Towards the Quantitative Prediction of the Phase Behavior of Polymer Solutions by Computer Simulation. *Macromolecular Symposia* **278,** 1–9 (2009).

77. Müller, E. A. & Jackson, G. Force-Field Parameters from the SAFT-γ Equation of State for Use in Coarse-Grained Molecular Simulations. *Annual Review of Chemical and Biomolecular Engineering* **5,** 405–427 (2014).

78. Lyubartsev, A. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52.** 00390, 3730–3737 (1995).

79. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128,** 244114 (2008).

80. Lyubartsev, A., Mirzoev, A., Chen, L. & Laaksonen, A. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss.* **144,** 43–56 (2010).

81. Motevaselian, M. H., Mashayak, S. Y. & Aluru, N. R. Extended coarse-grained dipole model for polar liquids: Application to bulk and confined water. *Physical Review E* **98** (2018).

82. Rosenberger, D. & Van der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Physical Chemistry Chemical Physics* **20,** 6617–6628 (2018).

83. Izvekov, S. & Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *The Journal of Physical Chemistry B* **109,** 2469–2473 (2005).

84. Izvekov, S. & Voth, G. A. Multiscale coarse graining of liquid-state systems. *The Journal of Chemical Physics* **123,** 134105 (2005).

85. Dama, J. F. *et al.* The Theory of Ultra-Coarse-Graining. 1. General Principles. *Journal of Chemical Theory and Computation* **9,** 2466–2480 (2013).

86. Davtyan, A., Dama, J. F., Sinitskiy, A. V. & Voth, G. A. The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation. *Journal of Chemical Theory and Computation* **10,** 5265–5275 (2014).

87. Dama, J. F., Jin, J. & Voth, G. A. The Theory of Ultra-Coarse-Graining. 3. Coarse-Grained Sites with Rapid Local Equilibrium of Internal States. *Journal of Chemical Theory and Computation* **13,** 1010–1022 (2017).

88. Ardham, V. R., Deichmann, G., Van der Vegt, N. F. A. & Leroy, F. Solid-liquid work of adhesion of coarse-grained models of n-hexane on graphene layers derived from the conditional reversible work method. *J. Chem. Phys.* **143,** 243135 (2015).

89. Nordholm, S. & Zwanzig, R. A systematic derivation of exact generalized Brownian motion theory. *Journal of Statistical Physics* **13,** 347–371 (1975).

List of Publications

The following four chapters, which constitute the results of this thesis have been published before in the following academic publications:

- **G. Deichmann**; M. Dallavalle; D. Rosenberger; N. F. A. van der Vegt, Phase Equilibria Modeling with Systematically Coarse-Grained Models - A Comparative Study on State Point Transferability, *Journal of Physical Chemistry B*, **2019**, 123 (2), p. 504-515
  Reprinted with permission from the Journal of Physical Chemistry B. Copyright 2019 American Chemical Society.

- **G. Deichmann**; N. F. A. van der Vegt, Conditional Reversible Work Coarse-Grained Models of Molecular Liquids with Coulomb Electrostatics - a Proof of Concept Study on Weakly Polar Organic Molecules, *Journal of Chemical Theory and Computation*, **2017**, 13 (12), p. 6158-6166
  Reprinted with permission from the Journal of Chemical Theory and Computation. Copyright 2017 American Chemical Society.

- **G. Deichmann**; N. F. A. van der Vegt, Conditional Reversible Work Coarse-Grained Models with Explicit Electrostatics - An Application to Butylmethylimidazolium Ionic Liquids, *Journal of Chemical Theory and Computation*, **2019**, 15 (2), p. 1187-1198
  Reprinted with permission from the Journal of Chemical Theory and Computation. Copyright 2019 American Chemical Society.

- **G. Deichmann**; N. F. A. van der Vegt, Bottom-up approach to represent dynamic properties in coarse-grained molecular simulations, *Journal of Chemical Physics*, **2018**, 149, p. 244114
  Reproduced from the Journal of Chemical Physics, with the permission of AIP publishing.

3 Phase Equilibria Modeling with Systematically Coarse-Grained Models - A Comparative Study on State Point Transferability

Abstract

Coarse-grained models for soft matter systems can be parameterized using a variety of approaches. In systematic coarse-graining procedures effective interactions are calculated using a reference fine-grained simulation. The degree to which thermodynamical and structural properties of the reference system are reproduced depends critically on the coarse-graining method used to parameterize the model. In this article a comparative study is presented on the capability of different coarse-grained models to reproduce vapor-liquid equilibrium thermodynamics of hexane and perfluorohexane systems. Coarse-grained models are developed using coarse-graining methods based on the reproduction of (*a*) structure and (*b*) coupling free energies. Although the structure-based models show an overestimation of the pressure and therefore perform relatively poorly in reproducing the vapor-liquid thermodynamics, it is shown that methods based on coupling free energies, in particular the conditional reversible work method, are capable of better reproducing the vapor-liquid phase diagram of the united-atom reference simulations, while reproducing the liquid structure almost as accurately as the structure-based coarse-grained model. This illustrates the state point transferability of the interaction potentials calculated using these methods as the resulting coarse-grained model (which has been parameterized at 300 K) is capable of accurately modeling the thermodynamic properties up to the critical point.

3.1 Introduction

Molecular simulations are an important tool for soft matter research.[1–3] While atomistically resolved models give us a very detailed picture, we can study systems on larger time and length scales when using a coarse-grained (CG) model. In these models, groups of atoms are represented by single sites, thus reducing the number of degrees of freedom in the system. The generation of a CG model from an underlying fine-grained (usually atomistic) model is called bottom-up coarse-graining. To generate a CG model in this manner a mapping scheme is constructed that relates the fine-grained (FG) to the CG representation. The key objective of the bottom-up coarse-graining procedure is then the parametrization of potentials for the interactions between the sites; various methods have been proposed to tackle this challenge.[2, 4–11]

The quality of a CG model can be evaluated, first, by the degree to which it is able to reproduce the properties of the reference model at the state point at which the model has been parameterized (representability). The second criterion for the quality of a CG model is the capability of reproducing the reference behavior at other state points (transferability) also, i.e. at a different temperature, a different phase, or a different composition in a mixture of components. The degree to which the CG models are representable and/or transferable critically depends on the method that is used to parameterize its (nonbonded) interaction potentials.

Structure-based coarse-graining methods have been developed with the intent of reproducing the bulk liquid structure of a reference system. The thermodynamic properties (e.g. pressure) can be reproduced by these methods by including additional parametrization targets into the coarse-graining procedure [5, 12, 13] or by multi-state parametrization of the CG potential.[7, 14–21] Another approach to coarse-graining is the attempt to reproduce the multi-body potential of mean force (PMF) of the CG system by means of reproducing the force acting on the mapped CG sites in the fine-grained representation.[6, 7] It has been shown that the multiscale coarse-graining (MS-CG) method, which is the most prominent among the force-based methods, will result in an optimal solution for a given state point.[6] With the aim

of parametrizing potentials which are transferable between state points, methods have been formulated to approximate the force-matching potentials by the use of coupling free energies between the atoms contained within the CG sites. This coupling free energy can be thought of as the free energy difference of the process of "switching-on" the interaction between pairs of atoms in the model.

These methods (EF-CG[8] and CRW[10]) can also be described as mean-field approximations to the exact force matching potentials obtained through the multiscale coarse-graining procedure. Thermodynamic properties can in many cases be reproduced accurately and in a transferable manner using the potentials parameterized using such coupling free energy methods.[8, 22–25] In addition, the models generated with these methods are often more transferable than structure-based models, because their interaction potentials do not include higher-order correlations that are projected onto pair interactions. The inclusion of information of higher-order correlations into the force-field can be beneficial to the transferability if performed correctly.[21] However, in the case of IBI or IMC, the higher-order effects on the pair correlation are projected onto the pair potential in a nontransferable manner, minimizing the CG entropy difference.[26, 27] In this case, neglecting the higher-order contributions to the pair potential altogether, as done in the CRW and EF-CG approaches, may be a better choice with regard to the transferability. The so-gained improvement in terms of the thermodynamic transferability comes at the cost of a slightly worse reproduction of the reference FG liquid structure.

In this work we will parameterize CG models based on the TraPPE-UA model for hexane and perfluorohexane using two coupling-free-energy-based methods, Conditional Reversible Work (CRW)[10, 28] and Effective-Force Coarse-Graining (EF-CG)[8], and one structure-based method, Inverse Monte Carlo (IMC)[4].

Our choice of hexane and perfluorohexane as test system is motivated, first, by the fact that these molecules show a miscibility gap at temperatures slightly below 300 K.[29] Second, the mapping of six heavy atoms to three CG sites, as used in this work, has been observed in various studies and has been used to generate multiple CG models of good quality.[10, 30, 31]

We will evaluate the transferability of the CG models (which are all parameterized at 300 K) by comparing their capability of reproducing the vapor-liquid coexistence curve (VLCC) up to the critical point. Further, we will investigate mixing properties of a binary liquid mixture of the two molecules. It has been shown that the expressions for thermodynamic observables in CG representations may not necessarily equal to those in the FG system and that therefore corrections to the CG results may be necessary to compare these values.[32] Still, the results presented in this article are calculated in an identical manner for both representations. Although many properties of the atomistic reference model are reproduced well, we cannot rule out completely the possibility of errors because of the representability problem. An estimation of these errors, however, would go beyond the scope of this work, even more so as the aim of this article is to compare different coarse-graining methods rather than to parameterize new models with perfect representability.

Various CG force fields exist in the literature that have been parameterized with the aim of reproducing the vapor-liquid equilibrium in a top-down manner.[33–35] These models perform well in reproducing thermodynamical properties but often lack predictive quality in terms of the liquid structure. Bottom-up models can, in many cases, describe the structure and thermodynamic properties well but to the authors' knowledge no studies have so far been attempted to investigate their quality in reproducing VLCCs.

The remainder of the article is structured as follows: in section 3.2 we will shortly summarize the coarse-graining methods used; the details of the simulations performed are listed in section 3.3; in section 3.4 the results are presented; and the article is concluded by a discussion of the results in section 3.5.

3.2 Methods

3.2.1 Coarse-Graining Methods

Conditional Reversible Work

The conditional reversible work (CRW) method[10] solves the coarse-graining problem by computing the coupling free energies between two groups of atoms by means of a thermodynamic cycle. A sequence of atomistic simulations of two molecules set at a given distance is performed for the parametrization of the potential. For each step of the atomistic trajectory the force between the molecules is sampled along the unit vector **e** parallel to the distance vector of the sites of interest. To obtain the reversible work between the CG sites we integrate the averaged constraint force over the coordinate r:

$$W(r) = -\int_{\infty}^{r} < \mathbf{F} \cdot \mathbf{e} > (r')dr' \tag{3.1}$$

This procedure is repeated twice, switching on and off the interactions of the atoms belonging to the two CG beads. The reversible work expended upon introducing the nonbonded interaction is:

$$U(r) = W(r) - W_{excl}(r) \tag{3.2}$$

Under the premise of having pairwise interactions, but without any assumptions on a particular form of the potential, CRW returns the CG interaction between two sites. The resulting potential is derived in a specific bonded (intramolecular) chemical environment. The surroundings affect the sampling of the forces between the molecules and, therefore, the potential is derived under the *condition* of having the beads embedded in their respective molecules. In the present case, the sampling has been performed in vacuum.

Effective-Force Coarse-Graining

The effective-force coarse-graining (EF-CG) method was formulated by Wang et al. [8] as a variant of the force-matching approach to systematic coarse-graining. Instead of optimizing a CG force field such that the atomistic *total* force is reproduced, the EF-CG method aims at a reproduction of the mean *pair* force between the CG sites. The rationale behind this parametrization is that pair interactions parameterized using global optimization schemes (such as force-matching) will always contain contributions from multi-body correlations and although these may be necessary to achieve an accurate representation of a single state point, the state point transferability suffers from their incorporation.[8] In an EF-CG model a pair of CG sites i and j containing atoms a and b interacts with the mean force F_{ij}, which is calculated by summing the atomistic pair force between the constituting atoms $\mathbf{F}_{ab}$ and projecting the result on the ij-distance vector $\mathbf{r}_{ij}$:

$$F_{ij}(r) = \left\langle \left(\sum_{a \in i} \sum_{b \in j} \mathbf{F}_{ab} \right) \cdot \frac{\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} \right\rangle_r , \tag{3.3}$$

where $\langle \ldots \rangle_r$ denotes a conditional average at $r_{ij} = r$.

The interaction potential of the CG model can be obtained by integrating the effective pair force:

$$U_{\mathrm{EF-CG}}(r) = -\int_{\infty}^{r} F_{ij}(r')dr' \tag{3.4}$$

It can be shown that the EF-CG interaction potentials are similar, but not exactly equal, to those calculated with the CRW method (see section 2.4.2 for more detail on the similarity between the two methods). A difference is mainly introduced by the fact, that the EF-CG method is usually applied to reference trajectories sampled in a liquid environment, as opposed to the vacuum sampling performed in the parametrization of most CRW models.

Inverse Monte Carlo

Inverse Monte Carlo (IMC), developed by Lyubartsev and Laaksonen[4], is an exact Newton inversion technique to derive effective pair potentials, which accurately reproduce the structure of the underlying fine grained (FG) system in the coarse grained (CG) configuration space. In IMC, the nonbonded pair interactions are derived based on PMF, U_{ij}^0, acting between two particles i and j. The PMF is obtained from the radial distribution function (RDF) between two particles i and j in the mapped FG configuration space, $g_{ij}^0(r)$, according to:

$$U_{ij}^0(r) = -k_\mathrm{B} T \ln g_{ij}^0(r), \tag{3.5}$$

where k_B is the Boltzmann constant T is the temperature of the system.

In many cases the PMF fails to accurately describe the pair distribution between i and j in the CG configuration space, because of the pertinence of higher-order correlations. Therefore, the PMF is updated n times, until the difference between the RDF of the FG and CG system is minimized:

$$U_{ij}^n = U_{ij}^{n-1}(r) - \Delta U_{ij} \tag{3.6}$$

$$\Delta U_{ij} = \mathbf{A}^{-1}(g_{ij}^{n-1}(r) - g_{ij}^0(r)), \tag{3.7}$$

where $\mathbf{A}$ is the Jacobi matrix,

$$\mathbf{A} = \frac{\partial g_{ij}}{\partial U_{ij}} \tag{3.8}$$

which can be expressed in terms of particle number fluctuations.[36]

The IMC framework provides an exact way to derive structure-based CG models, but underlies the typical constraints of iterative methods to solve an inverse problem, which is the dependence on a well-conditioned initial guess, numerical stability and sufficient sampling.[17]

3.2.2 Gibbs Ensemble Monte Carlo

Two main limitations can be encountered when directly simulating coexisting phases: the presence of finite size interfacial effects and difficulties in equilibrating the system associated to the concentration fluctuations.[37] When two or more phases coexist in a single small-sized box, a large percentage of molecules are located in the interfacial region. The effect of the interface on the thermodynamic properties tends to be over-represented. The system is interface-dominated, impeding a reliable prediction of bulk macroscopic properties, such as the critical temperature. In case of mixtures the coexisting phases are hard to equilibrate as lengthy relaxation times for the concentration fluctuations are to be expected. To overcome the limitations in equilibrating coexisting phases using molecular dynamics, Gibbs ensemble Monte Carlo (GE-MC) simulations can be performed.[38] In GE-MC the interface is absent and two or more separate boxes are simulated. The two boxes are coupled such that (at equilibrium) temperature, pressure and chemical potential in the two regions are the same, which is the condition of phase coexistence. Across the twin boxes a number of Monte Carlo moves is performed to sample different configurations of the boxes and to exchange particles and volume between the boxes.

Below the critical point, GE-MC simulations can be set up with two boxes representing liquid and vapor phases, respectively. From such a simulation saturation vapor pressures as well as vapor and liquid densities can be calculated.

3.2.3 Calculation of Mixing Properties

Mixing properties are calculated in this work to characterize binary liquid mixtures at different compositions to judge the transferability of the CG potentials with respect to the composition of the system. We calculate the excess volume of mixing $\Delta_{\text{mix}}V$ and excess enthalpy of mixing $\Delta_{\text{mix}}H$ for a binary mixture, using simulation results at a given composition (mole fraction) x:

$$\Delta_{\text{mix}}V(x) = V(x) - [(1-x)V(0) + xV(1)] \tag{3.9}$$

$$\Delta_{\text{mix}}H(x) = H(x) - [(1-x)H(0) + xH(1)] \tag{3.10}$$

Another thermodynamic quantity that we compute in this work is the excess free enthalpy of mixing $\Delta_{\text{mix}}G$, which can be calculated using Kirkwood-Buff (KB) theory.[39] The KB theory allows the calculation of thermodynamic properties from structural correlations in the liquid. The central quantity in this theory is the KB integral G_{ij} for components i and j, which can be calculated from the radial distribution function $g_{ij}(r)$:

$$G_{ij} = \int_0^\infty 4\pi r^2 (g_{ij}(r) - 1) dr \tag{3.11}$$

The definition above, strictly applies only to infinitely large systems in a grand canonical ensemble; for finite size systems at constant N, correction terms have to be included.[40, 41] In this work, Ganguly's RDF correction is used for the calculation of G_{ij} [42], which replaces the $g_{ij}(r)$ in the integral by a corrected term $g'_{ij}(r)$:

$$g'_{ij}(r) = g_{ij}(r) \frac{N_j\left(1 - \frac{4\pi r^3}{3V}\right)}{N_j\left(1 - \frac{4\pi r^3}{3V}\right) - \Delta N_{ij}(r) - \delta_{ij}}, \tag{3.12}$$

where N_j denotes the total number of j molecules, V denotes the box volume, $\Delta N_{ij}(r) = N_j G_{ij}(r)/V$ is the uncorrected excess number at distance r, and δ_{ij} denotes the Kronecker delta.

For a binary mixture $\partial^2 \Delta_{\text{mix}}G / \partial x_1^2$ is related to the KB integral by:[43]

$$\frac{\partial^2 \Delta_{\text{mix}}G}{\partial x_1^2} = \frac{k_B T \rho}{x_1 x_2 [\rho_1 + \rho_2 + \rho_1 \rho_2 (G_{11} + G_{22} - 2G_{12})]}, \tag{3.13}$$

where $\rho = \rho_1 + \rho_2$ denotes the total number density.

From the second derivative $\Delta_{\text{mix}}G$ can be computed using the iterative shooting method by Galata et al.[43] Here, a value for $\frac{\partial \Delta_{\text{mix}}G}{\partial x}(x=0)$ is guessed and $\Delta_{\text{mix}}G(x)$ is calculated as:

$$\Delta_{\text{mix}}G(\Delta x) = \Delta_{\text{mix}}G(0) + \Delta x \frac{\partial \Delta_{\text{mix}}G}{\partial x}(0) \tag{3.14}$$

$$\Delta_{\text{mix}}G(x_i + \Delta x) = \Delta x^2 \frac{\partial^2 \Delta_{\text{mix}}G}{\partial x_1^2}(x_i) + 2\Delta_{\text{mix}}G(x_i) - \Delta_{\text{mix}}G(x_i - \Delta x), \tag{3.15}$$

where $\Delta x = 0.05$ is a finite concentration increment. The initial value for $\frac{\partial \Delta_{\text{mix}}G}{\partial x}(x=0)$ is then modified using a Newton-Raphson scheme such that the resulting $\Delta_{\text{mix}}G(1) = 0$.

Eq. 3.13 is only valid in the stable regions of the phase diagram, i.e. for concentrations where $\partial^2 \Delta_{\text{mix}}G / \partial x^2 > 0$. Therefore we will not be able to reconstruct a liquid-liquid phase diagram from RDF data, even from trajectories where phase separation can be observed. Instead, we will compare $\Delta_{\text{mix}}G$ predicted by different CG models at 300 K for a (homogeneous) binary mixture of hexane and perfluorohexane.

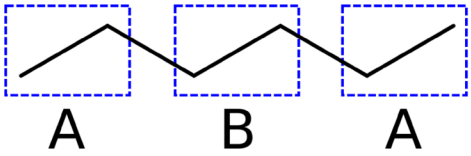


Figure 3.1: Mapping scheme used for hexane with three CG sites (A-B-A). An analogous mapping is used for perfluorohexane (C-D-C).

3.3 Computational Details

3.3.1 Model Systems

The molecules considered in this study are (*n*)-hexane and (*n*-)perfluorohexane. For the FG representation the TraPPE united atom model is chosen[44, 45], which has been parameterized to correctly predict the vapor-liquid phase equilibria of the pure components. In this model, all nonbonded interactions are cut-off at 1.4 nm and long-range tail corrections are applied for Lennard-Jones energy and pressure.

In the CG model each molecule is represented by three sites, containing two terminal and one central CH_x/CF_x-groups respectively. An illustration of the mapping scheme for hexane is shown in fig. 3.1. As depicted, each hexane molecule consists of two (terminal) A-sites and one (central) B-sites. The perfluorohexane molecules are modeled in an analogous manner using sites labeled C (terminal) and D (central).

3.3.2 Coarse-Grained Models

The bonded degrees of freedom are modeled identically in all CG models. Bond length and bond angle distributions $\rho(r_\mathrm{b})$ and $\rho(\theta)$ are gathered from an atomistic MD simulation of one molecule in vacuo. In this simulation the time step is 1 fs and the overall simulation length is 100 ns. The temperature of 300 K is maintained using a Langevin dynamics thermostat[46] with a time constant of 1 ps. The CG potentials U are calculated from these distributions by Boltzmann inversion:

$$U_\mathrm{bond}(r) = -k_\mathrm{B}T \ln\left(\frac{\rho(r)}{r^2}\right) + C_\mathrm{bond} \tag{3.16}$$

$$U_\mathrm{angle}(\theta) = -k_\mathrm{B}T \ln\left(\frac{\rho(\theta)}{\sin(\theta)}\right) + C_\mathrm{angle}, \tag{3.17}$$

where C denotes a constant that is chosen such that the minimum of the respective U is zero.

The CRW model is parameterized in vacuo using the Monte-Carlo sampling algorithm described in ref. [47] to sample the forces required for the calculation of W and W_excl (at 300 K). The range of distances r considered has been 0.32-1.50 nm with a distance increment of 0.02 nm and each sampled trajectory has been 10^7 trial steps in length.

The EF-CG model is parameterized using a mixture with a composition of 150 molecules of hexane and perfluorohexane each. This simulation has a length of 5 ns with a time step of 1 fs and is performed in a NpT-ensemble at 300 K and 1 bar using GROMACS 4.6.7.[48] The Berendsen thermostat and barostat[49] are used to maintain the temperature and pressure with time constants of 0.5 ps and 1 ps, respectively. The EF-CG potential parametrization is performed by post-processing the trajectory generated from this simulation, calculating the mean nonbonded pair force between CG mapping points at a given distance. The resulting potentials are not very sensitive to the composition of the liquid phase (see fig. 3.2) and therefore the potentials obtained from the 1:1 mixture can also be used simulations of the pure components.

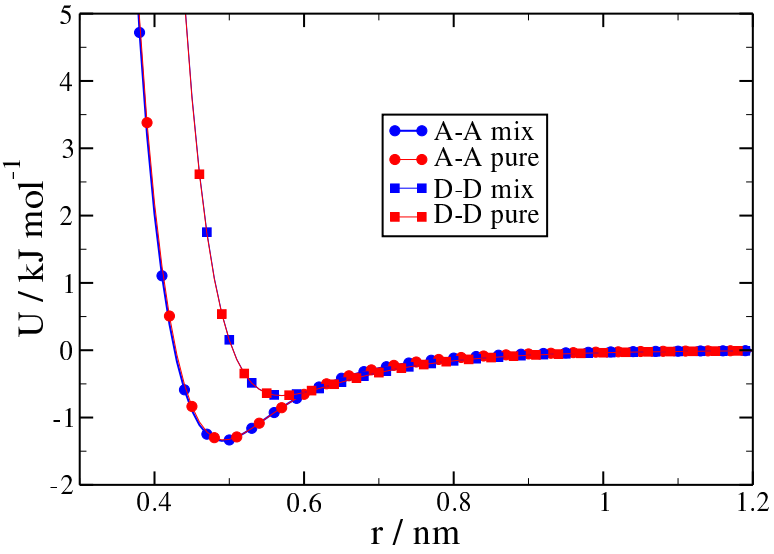


Figure 3.2: The nonbonded potentials calculated through the EF-CG method are not sensitive to composition. The graph compares the potential curves for the A-A (circles) and D-D (squares) interactions calculated using EF-CG with trajectories sampled in the 1:1 mixture (blue curves) and in the pure liquid (red curves), respectively.

The nonbonded IMC potentials have been obtained with the aid of the VOTCA toolkit (version 1.4)[30, 50] in combination with the Gromacs 2016-4 simulation package.[51, 52] On the basis of liquid phase simulations of pure FG hexane, pure FG perfluorohexane, and the 1:1 mixture the target RDFs for the different site-site interactions have been obtained. The different RDFs have been evaluated according to table 3.1 with a bin size of 0.01 nm. Further, the PMFs obtained from the different target RDFs have been used as an initial guess for the iterative procedure. At each iteration a short MD simulation of 2 ns has been performed with a timestep of 1 fs. As an integrator the leap-frog stochastic dynamics integrator[53] has been applied with an inverse friction constant of 1 ps. The temperature was kept constant at 300 K. For the pure hexane system, we performed 125 iterations in total, including 30 iterations of IBI to get a better initial guess than the PMF. The same applies for perfluorohexane, but here we only performed 10 preliminary steps of IBI. For the binary mixture the CRW potentials served as an initial guess for the 31 IMC iterations, where each intra- and inter-species interaction got updated every third iteration.

The potentials parameterized from the mixed system are different from those calculated for the pure components (see fig. 3.3). For calculating the liquid-vapor equilibria the potential parameterized in the pure liquids are used. For the simulations in the liquid mixtures the potentials parameterized from the 1:1 mixture are used to assess their transferability with respect to the composition.

For the subsequent simulations with the CG models, potential tables are set up using spline interpolation with a distance increment of 0.001 nm. The cutoff length in the CG simulations is 1.5 nm and no long-range corrections are used in the CG simulations.

3.3.3 Liquid Phase Simulations

For the calculation of liquid structure and mixing properties, simulations are performed for mixtures of hexane and perfluorohexane in a simulation box containing 500 molecules using the GROMACS 4.6.7 software package. An NpT ensemble at 300 K and 1 bar is simulated using the Nosé-Hoover thermostat[54] and Parrinello-Rahman barostat[55] with time constants of 0.5 ps and 2.0 ps. Simulations with

Table 3.1: Inner and outer cutoff values used in the IMC parametrization for all interaction pairs

System	interaction type	r_min	r_max
hexane	A-A	0.32	1.5
•	A-B	0.33	1.5
•	B-B	0.33	1.5
perfluorohexane	C-C	0.38	1.5
•	C-D	0.38	1.5
•	D-D	0.39	1.5
mixture	A-C	0.37	1.5
•	A-D	0.38	1.5
•	B-C	0.38	1.5
•	B-D	0.39	1.5

the IMC model are conducted at NVT conditions at 300 K and the simulation time is 10 ns with a simulation step of 2 ps. In the atomistic simulations all bond distances are constrained using the LINCS algorithm.[56] The composition is varied from pure hexane ($x = 0$) to pure perfluorohexane ($x = 1$) in increments of 5 mol-%.

3.3.4 Simulations of the Vapor-Liquid Phase Equilibrium

Gibbs ensemble Monte Carlo simulations are performed with all models using the MCCCS Towhee[57] simulation code, which was modified by us to allow for the simulation of tabulated bonds and angles. NVT simulations are performed with two simulation boxes and a total number of 300 molecules at ≈10 different temperatures below the critical temperature with an increment of 10 K. After 5000 cycles of equilibration a production run with 10000 cycles (50000 for perfluorohexane) is performed. Every cycle consists of 300 of the following trial moves (listed with their respective probability): volume change (0.2%), exchange of particles between boxes (9.8%), configurational bias regrowth[58] (23%), rotation (33%), and displacement of a molecule within the box (34%).

3.4 Results and Discussion

3.4.1 Coarse-Grained Models

The nonbonded potentials obtained using different coarse-graining methods are shown in figure 3.3. The potentials calculated with IMC deviate from the other two sets of potentials, whereas the potentials parameterized using CRW and EF-CG are very similar in shape.

In fig. 3.4 the potentials for the bond distance and bond angle of the two molecules are shown. Owing to the different torsional potentials in the atomistic model, the bond and angle distributions of both molecules differ significantly. This difference is reflected by the different structure, especially of the angular potentials.

3.4.2 Structure of Liquid Systems

The radial distribution functions $g(r)$ calculated from the pure component liquid simulations are shown in fig. 3.5. These RDFs serve as a validation of the representability of the CG models. As expected, the IMC model performs best in reproducing the atomistic structure of the system: it quantitatively reproduces the $g(r)$ of the FG system. This is an expected result as the aim of iterative structure-based methods like IMC is to achieve exactly this aim.

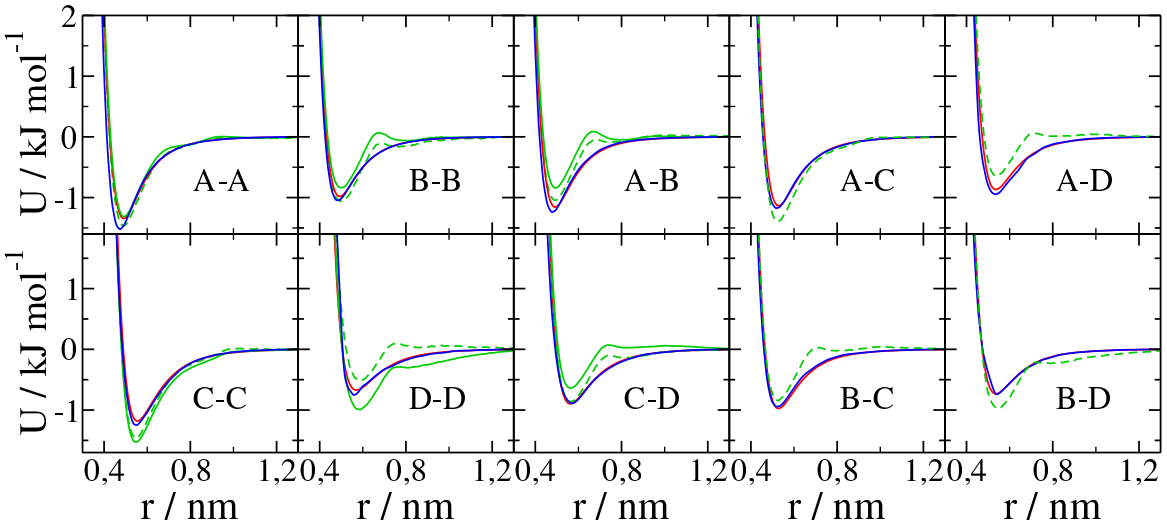


Figure 3.3: Nonbonded CG potentials for all site-site pairs parameterized using EF-CG (red), CRW (blue), and IMC (green) methods. Two sets of potentials parameterized with IMC are shown which were derived in pure liquids (solid lines) and a 1:1 binary mixture (dashed lines), respectively.

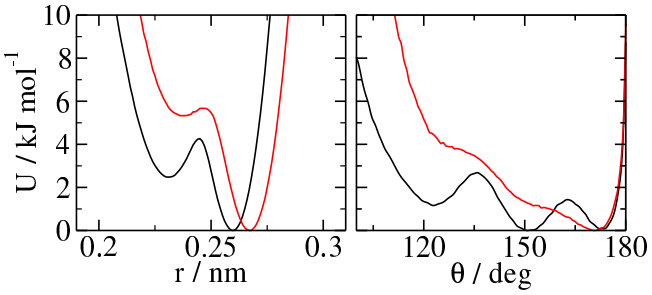


Figure 3.4: Bonded CG potentials for hexane (black) and perfluorohexane (red) bond distances (right half) and angles (left half).

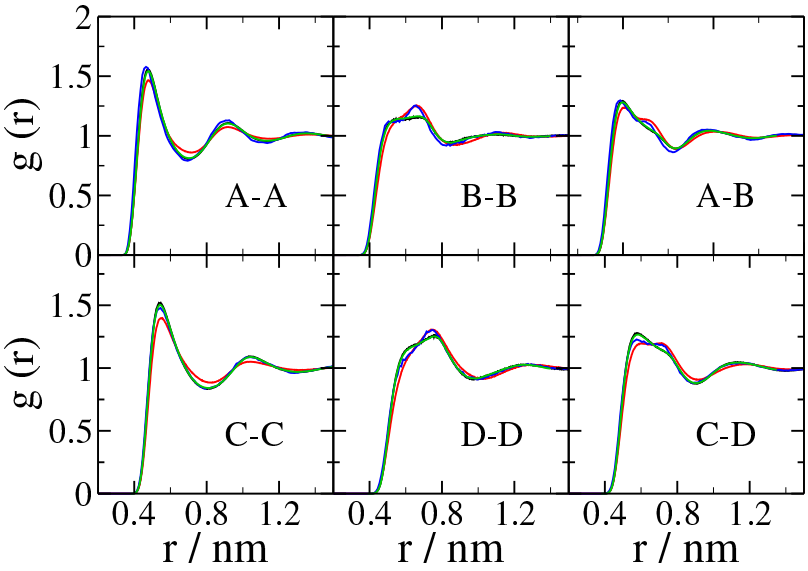


Figure 3.5: Radial distribution functions from pure component simulations using the TraPPE-UA (black), EF-CG (red), CRW(blue), and IMC (green) models.

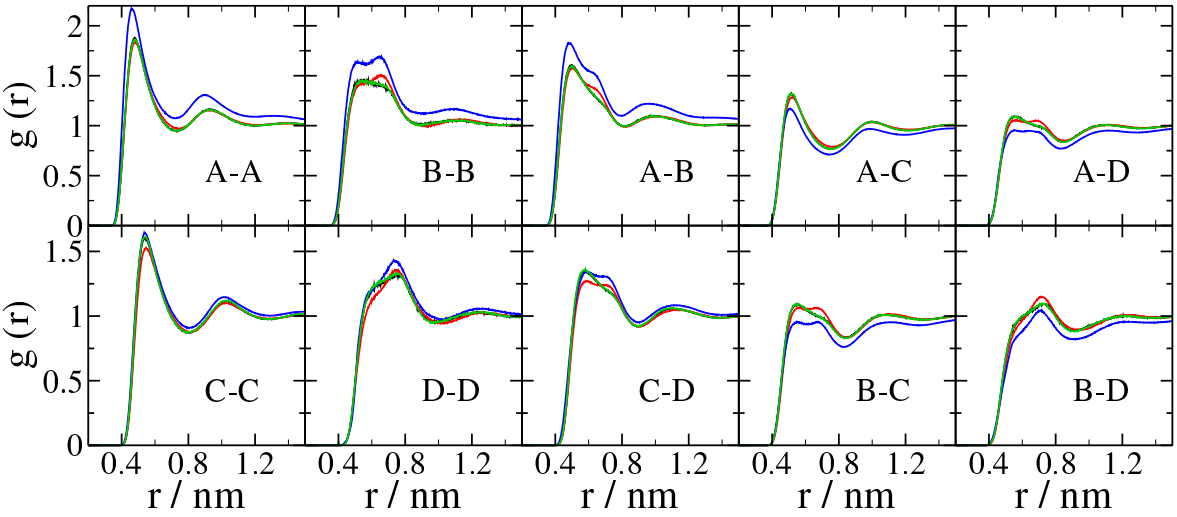


Figure 3.6: Radial distribution functions from simulations of a 1:1 binary mixture ($x = 0.5$) using the TraPPE-UA (black), EF-CG (red), CRW(blue), and IMC (green) models.

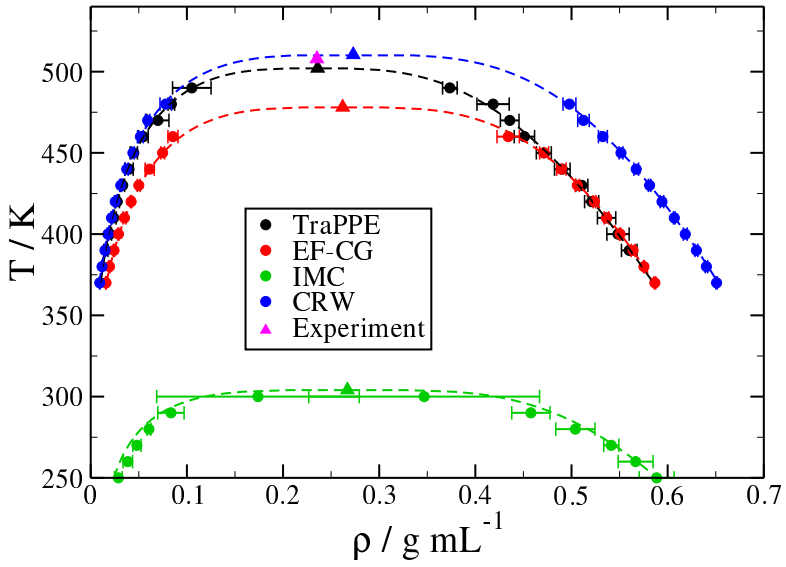


Figure 3.7: Vapor-liquid coexistence curves of hexane. Dashed lines denote fit to obtain critical point. Triangles denote the critical point.

The other two models do not reproduce the liquid structure to the same degree as the IMC model, but still show a good agreement with the united-atom reference date. This is an important result because both models are not parameterized using any structural correlation as input. Especially for the pairs A-A and C-C, the CRW model performs slightly better than the EF-CG model with respect to the reproduction of $g(r)$ for all CG site pairs. This is a somewhat remarkable result, since the CRW potentials have been parameterized using only two molecules in vacuo, whereas the EF-CG model has been parameterized from liquid phase simulations. However, the deviations between the RDF from simulations with these two models are rather small.

The structure of a liquid system containing 50 mol-% hexane and perfluorohexane each is shown in figure 3.6. Here again, the IMC model reproduces the reference structure to a quantitative degree and the EF-CG model reproduces the essential characteristics of the RDF shape well, without achieving a quantitative reproduction. The results obtained with the CRW model look rather different. Although the shape of the RDF curves is in line with those of the EF-CG model, an overestimation of the hexane-hexane pairs and an underestimation of the hexane-perfluorohexane pairs can be observed. This suggests that this system is close to phase separation in the liquid phase, which would suggest an overestimation of the upper critical solution temperature by the CRW model. The RDFs obtained using the EF-CG model do not show signs of clustering of the two components. We would like to recall here that the EF-CG model has been obtained from sampling the liquid state. It is possible that the more realistic phase space sampling in the EF-CG parametrization benefits the correct reproduction of the liquid-liquid phase equilibrium.

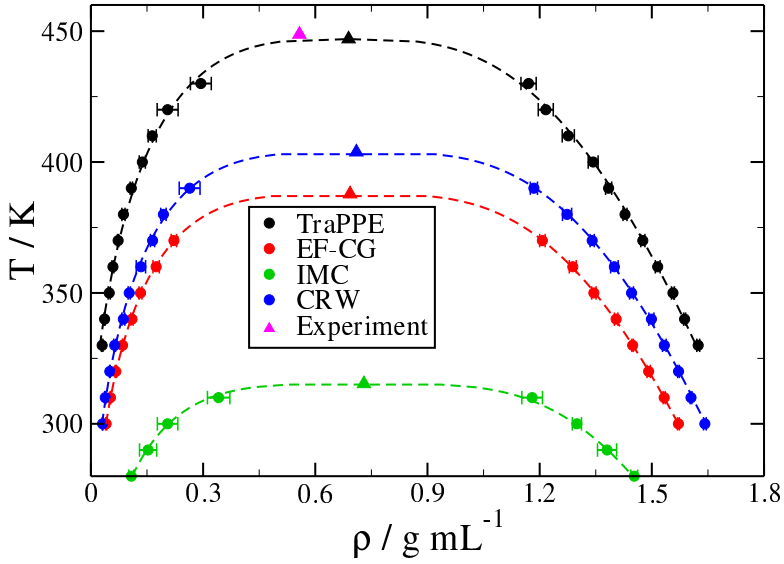


Figure 3.8: Vapor-liquid coexistence curves of perfluorohexane. Dashed lines denote fit to obtain critical point. Triangles denote the critical point.

Table 3.2: Critical properties and fit parameters

system		T_c / K	ρ_c / g mL^{-1}	A / 10^{-4} g mL^{-1} K^{-1}	B g mL^{-1} K^{-c}	c
hexane	TraPPE	502	0.24	4.67	0.124	0.314
	CRW	510	0.27	4.12	0.162	0.279
	EF-CG	478	0.26	3.63	0.158	0.274
	IMC	304	0.27	7.31	0.215	0.242
	Exp.[59]	507.9	0.235			
perfluorohexane	TraPPE	447	0.69	11.3	0.366	0.309
	CRW	404	0.71	12.2	0.446	0.248
	EF-CG	388	0.69	12.8	0.446	0.276
	IMC	315	0.73	14.1	0.73	0.248
	Exp.[60]	448.8	0.558			

3.4.3 Vapor-Liquid Equilibrium

The vapor-liquid coexistence curves (VLCC) obtained from the GEMC simulations are shown in figures 3.7 and 3.8 for hexane and perfluorohexane, respectively. The critical properties can be extrapolated from these curves using the law of rectilinear diameters:

$$\frac{\rho_l + \rho_g}{2} = \rho_c + A(T_c - T) \tag{3.18}$$

$$\rho_l - \rho_g = B(T_c - T)^c, \tag{3.19}$$

where ρ_l, ρ_g, and ρ_c are liquid, vapor, and critical densities respectively and T_c denotes the critical temperature. The resulting fitting parameters and critical properties are listed in table 3.2.

The degree with which the atomistic VLCCs are reproduced varies strongly depending on the CG model used. The IMC model predicts critical temperatures that are 200 K (hexane) and 140 K (perfluorohexane) below the reference value. This result is not too surprising, as structure-based models usually do not reproduce the pressure of the underlying atomistic reference well and are therefore not well suited for the prediction of vapor-liquid phase equilibria. The EF-CG model performs much better for both molecules, yielding a relative deviation of 5% for hexane and 13% for perfluorohexane. Finally, the CRW model performs best for both test systems. The critical temperature is slightly overestimated for hexane (<2%) and underestimated for perfluorohexane with a relative deviation of less than 10%.

This result provides evidence for the fact that models parameterized using the CRW method are capable of accurately describing the behavior of molecular liquids at the vapor-liquid interface.

We suspect that the reason for the good performance of the pair-force based CG methods is their good representation of the attractive part of the nonbonded pair forces. The EF-CG and CRW models have been shown to reproduce the vapor-liquid surface tension accurately[8, 10]. This means that they are transferable with respect to the nature of the surroundings (bulk liquid, interface or gas phase) and this may also be the reason for these models to perform well in reproducing other inter-phase properties. Both methods predict similar VLCCs for both systems. The detailed behavior regarding the critical point is well in line with the vapor pressure dependence on the temperature (fig. 3.9). A model which systematically overestimates the vapor pressure will, in turn, underestimate the VLCC.

3.4.4 Thermodynamics of Vapor-Liquid Equilibria

From the vapor and liquid coexisting densities in the VLCC, the free energy of transfer of one molecule from the vapor to the liquid phase ΔG_{trans} can be calculated using:

$$\Delta G_{\text{trans}} = -k_{\text{B}} T \ln\left(\frac{\rho_{\text{liq}}}{\rho_{\text{vap}}}\right)_{\text{eq}}. \tag{3.20}$$

Theoretically ΔG_{trans} corresponds to the reversible work of introducing the noncovalent interactions of a molecule in the liquid phase with all surrounding molecules ("self-solvation"). A CG model reproducing ΔG_{trans} can therefore be assumed to be good in predicting the correct liquid phase thermodynamics.[24]

In addition, the saturation pressure is obtained from the simulations, by calculating the virial of the simulation boxes. Its temperature dependence can be used to calculate the vaporization enthalpy ΔH_{vap} using the Clausius-Clapeyron equation:

$$\Delta H_{\text{vap}} = T \Delta V^{\text{m}} \frac{dp}{dT}, \tag{3.21}$$

where $\Delta V^{\text{m}} = V^{\text{m}}{}_{\text{vap}} - V^{\text{m}}{}_{\text{liq}}$ denotes the difference in molar volumina of the vapor and liquid phases.

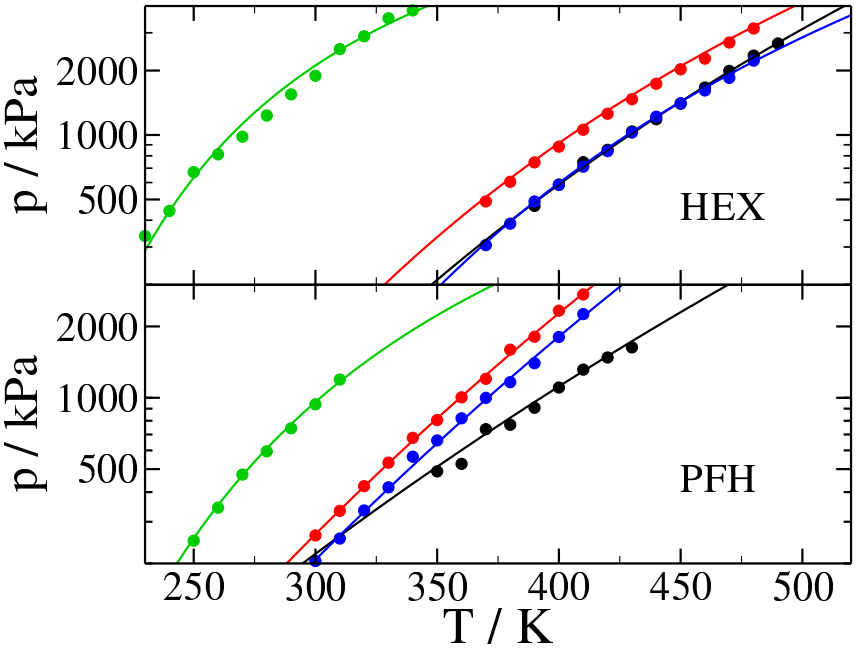


Figure 3.9: Calculated (circles) vapor pressures and fitting function (solid lines) for hexane (top) and perfluorohexane(bottom) from simulations with TraPPE (black), CRW (blue), EF-CG (red), and IMC (green) models.

To facilitate the calculation of dp/dT, we first fit the calculated pressures using an Antoine-type fitting function with three parameters (A_p, B_p, and C_p):

$$\ln \frac{p}{1\,\mathrm{kPa}} = A_\mathrm{p} - \frac{B_\mathrm{p}}{T + C_\mathrm{p}}. \tag{3.22}$$

The vapor pressures and fitting curves are shown in fig. 3.9.

The resulting enthalpies of vaporization and free enthalpies of transfer are plotted as a function of the reduced temperature in fig. 3.10. For hexane, the CRW and EF-CG models quantitatively reproduce both quantities as predicted by the TraPPE reference model. The IMC model performs slightly less well, as it underestimates the absolute value of ΔG_trans while reproducing ΔH_vap well.

For perfluorohexane, ΔG_trans is reproduced well by all models (as in the case of hexane). In contrast, ΔH_vap is only reproduced well by the IMC and CRW models, and the EF-CG model drastically underestimates the reference values from the UA simulation. This is caused the underestimation ΔV^m near the critical point, which can be observed in fig. 3.8.

3.4.5 Thermodynamics of Liquid Mixtures

In addition to the thermodynamic data presented for the vapor-liquid equilibria of the two pure components, we present in the following simulation results of liquid mixtures of hexane and perfluorohexane. Mixtures of hexane and perfluorohexane show a miscibility gap below an upper critical solution temperature of about 295 K. Simulations of a liquid-liquid equilibrium using GE-MC are, in principle, possible but are very difficult to converge at high densities and low temperatures. Therefore we study the mixing thermodynamics at constant temperature of 300 K (above the upper critical solution temperature) to determine whether the CG models are transferable with respect to the composition of the system within the miscible region.

In figure 3.11 we compare the total volume V^m of the mixture and the excess volume $\Delta_\mathrm{mix}V$. Both models (EF-CG and CRW) reproduce the excess volume of the atomistic model rather accurately. Here the CRW model performs a little better, but the differences are not too large.

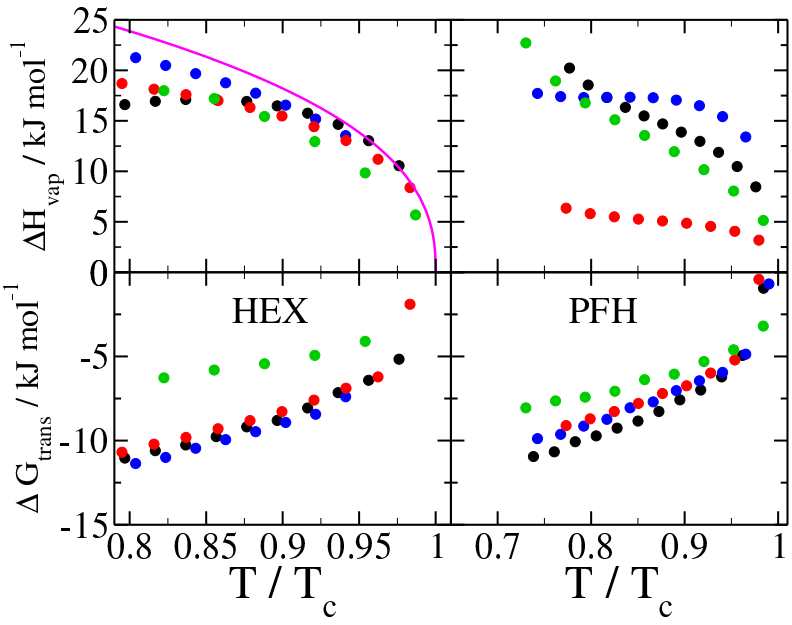


Figure 3.10: Vaporization enthalpies (top) and transfer free energies (bottom) calculated from the vapor-liquid equilibrium properties. The same coloring scheme as in fig. 3.9 is used. The magenta line depicts the experimental enthalpy of vaporization of hexane.[61]

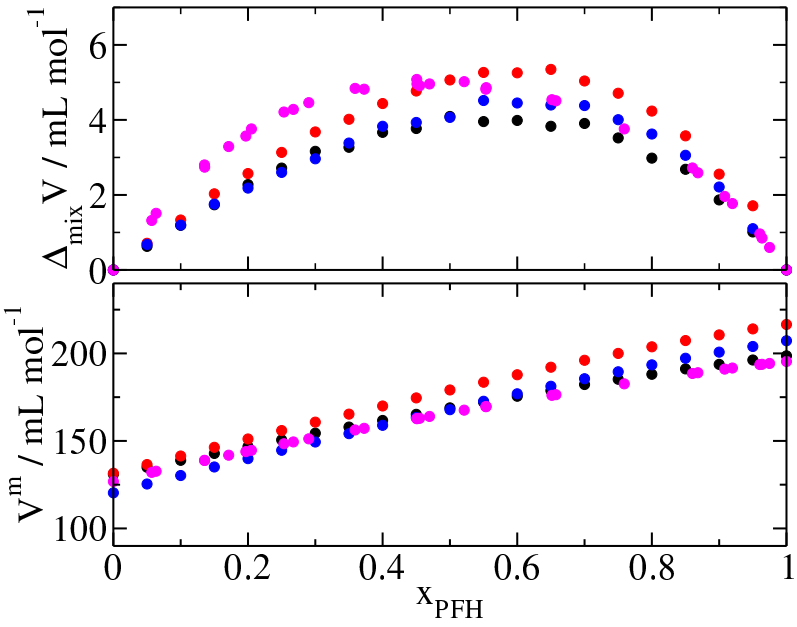


Figure 3.11: Excess volumes (upper panel) and molar volumes (lower panel) at 300 K. Simulations are performed using TraPPE-UA (black), EF-CG (red), and CRW (blue) models. Experimental data[62] (magenta) are shown for comparison.

The CG models are also similar in their reproduction of the molar volume, the EF-CG predicts higher volumina which leads to an overestimation of V^m at $x = 1$ and a good agreement to reference data at $x = 0$. The CRW model matches the UA molar volume at $x = 0.5$, and under- and overestimates V^m for pure hexane/perfluorohexane, respectively. Reproducing the molar volume requires an exact reproduction of the virial pressure which is, however, very sensitive to changes in the potential. On the one hand, this means that it is difficult to reproduce it exactly when not including it as a constraint in the potential parametrization. For CRW models, a slight mismatch in the prediction of V^m has been observed in earlier studies.[25, 63] On the other hand, a reproduction of the exact density can be obtained by adding slight modifications to the nonbonded potentials, which do not affect the remaining properties of the model to a large degree.[47]

The mixing enthalpy $\Delta_{\text{mix}}H$ and the free enthalpy of mixing $\Delta_{\text{mix}}G$ of the hexane-perfluorohexane mixtures are plotted in figure 3.12. In addition to the EF-CG and CRW models, we calculate these properties also from simulations with the IMC model parameterized at $x = 0.5$. These simulations are, in contrast to the simulations of the other models, carried out under NVT-conditions at the density of the atomistic model because the overestimation of the virial with the IMC model would lead to unrealistically low densities in an NpT simulation.

For $\Delta_{\text{mix}}H$ the trends observed in the atomistic model are reproduced by all CG models. The EF-CG and CRW models predict values close to the experimental and united-atom values whereas the IMC model underestimates $\Delta_{\text{mix}}H$.

A rather different behavior is observed for $\Delta_{\text{mix}}G$, an observable which can be used to estimate whether phase separation occurs in the liquid or not. In the case of phase separation we expect a low curvature of $\Delta_{\text{mix}}G(x)$ at concentrations around $x = 0.5$. This is actually observed for the values calculated from the simulation with the CRW model and this finding supports our conclusion from the RDFs of the mixture (fig. 3.6) that this system is actually very close to the critical solution temperature. All other curves present a clearly positive curvature, which means that the two components are miscible at any composition at the given temperature. Again, the EF-CG model is better than the IMC model in reproducing this thermodynamic property.

Also, the EF-CG model stands out as the model which reproduces best the values of $\Delta_{\text{mix}}H$ and $\Delta_{\text{mix}}G$ of the FG model. In addition to the results observed for the RDFs of the mixed systems, this can serve as an indication that the EF-CG model performs best, among the models studied, in reproducing the liquid-liquid phase behavior.

3.5 Conclusions and Outlook

In this article we have applied three different systematic coarse-graining methods to generate CG models for hexane and perfluorohexane. Molecular dynamics simulations have been conducted with these models in bulk liquid systems of the pure components and a mixture with $x_{\text{PFH}} = 0.5$. These simulations show that the shape of the CG site-site RDFs are well reproduced by all models. The RDFs of the CRW model show signs of a phase separation in the liquid phase at $x = 0.5$.

Further, Gibbs-ensemble MC simulations have been performed to calculate the vapor-liquid coexistence curves for the atomistic (TraPPE-UA) model and all CG models. The results show, that the CRW model performs best in reproducing the VLCC for both molecules. This confirms to us the transferability of the CRW models with regard to temperature and the phase transfer from the liquid to the gas phase. The EF-CG model performs less well in the reproduction of the VLCC and the liquid density although the interaction potentials are very similar to those of the CRW model.

We speculate that this mismatch may be caused by our use of bonded potentials parameterized from simulations of one molecule in vacuo. In the first publication on EF-CG Wang et al. used iteratively refined bonded potentials parameterized in a bulk liquid simulation.[8] These potentials may be better suited to complement the EF-CG nonbonded potentials in reproducing properties of the liquid.

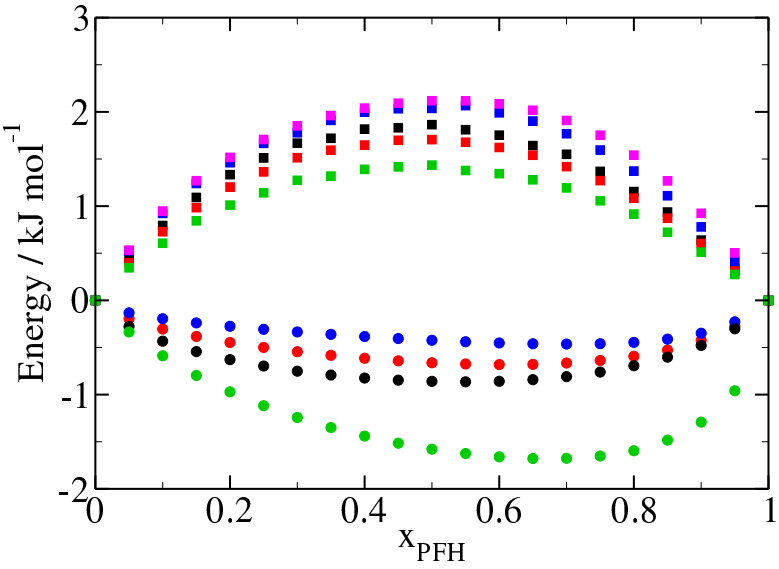


Figure 3.12: Excess enthalpy of mixing (squares) and excess free enthalpy of mixing (circles) at 300 K. Simulations are performed using TraPPE-UA (black), EF-CG (red), CRW (blue), and IMC (green) models. Experimental data[64] (magenta) is shown for comparison.

The IMC model accurately reproduces the liquid structure of the reference simulations, but reproduces the vapor-liquid equilibrium far less accurately than the two coupling-free-energy-based methods.

From the vapor-liquid equilibrium data and vapor pressures we have calculated vaporization enthalpies and free enthalpies of transfer as a function of the reduced temperature for all models. The results show that these properties are reproduced better by the EF-CG and CRW models than by the IMC model (with the notable exception of the EF-CG vaporization enthalpy of perfluorohexane), indicating that these force-based approaches perform better in the reproduction of the system's thermodynamics than the structure-based CG model.

Simulations of mixtures of the two components at varying composition are performed at 300 K. The results indicate that the EF-CG and CRW models show a better transferability with respect to the concentration than the IMC model. Moreover, we find that the CRW model predicts a critical solution temperature close to the simulated 300 K, as we can observe indications for phase separation of the liquid both from $\Delta_{\text{mix}}G$ and the site-site RDFs at $x = 0.5$.

We have demonstrated in this article that CG models parameterized through systematic (bottom-up) coarse-graining methods are capable of reproducing the thermodynamics of pure component vapor-liquid equilibria and liquid mixtures qualitatively, while also correctly reproducing the structure of the liquid phase. The quality of reproduction is better with coarse-graining methods that give a more realistic representation of the interactions in the system (CRW and EF-CG) than with the structure-based IMC models. To achieve a quantitative agreement in terms of the phase boundaries, slight modifications may be applied to the CG potentials generated with the CRW method. In a recent study we have shown, that a slight modification of the potentials can lead to significant increase in the representability with respect to the virial pressure, while leaving the liquid structure almost unchanged.[47]

The models parameterized in this work use a very low degree of coarse-graining (two sites being mapped to one). It might be interesting to evaluate in further studies how the results model quality with respect to the interphase thermodynamics changes with decreasing degree of resolution of the CG model. At some point, the nonbonded potential will no longer retain the rigid (atomistic) character observed for the CG models in this study and it is certainly interesting to see, whether or not the CG methods

can still predict the molar volume and thermodynamics correctly. For the case of the CRW method it has been shown by Dallavalle[63] that the use of lower resolution mappings is not detrimental to the quality (w.r.t. the molar volume) of the so-parameterized CG models. This indicates to us that the vapor-liquid equilibrium of an atomistic reference may also be well reproduced by coarser mappings.

3.6 References

1. Peter, C. & Kremer, K. Multiscale simulation of soft matter systems. *Faraday Discussions* **144.** 00053, 9 (2010).
2. Brini, E. *et al.* Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter* **9.** 00021, 2108 (2013).
3. Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* **139,** 090901 (2013).
4. Lyubartsev, A. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52.** 00390, 3730–3737 (1995).
5. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24.** 00428, 1624–1636 (2003).
6. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128,** 244114 (2008).
7. Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129,** 144108 (2008).
8. Wang, Y., Noid, W. G., Liu, P. & Voth, G. A. Effective force coarse-graining. *Physical Chemistry Chemical Physics* **11,** 2002 (2009).
9. Mullinax, J. W. & Noid, W. G. A Generalized-Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials †. *The Journal of Physical Chemistry C* **114,** 5661–5674 (2010).
10. Brini, E., Marcon, V. & Van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics* **13.** 00022, 10468 (2011).
11. McCarty, J., Clark, A. J., Copperman, J. & Guenza, M. G. An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *The Journal of Chemical Physics* **140,** 204913 (2014).
12. Wang, H., Junghans, C. & Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *The European Physical Journal E* **28,** 221–229 (2009).
13. Ganguly, P., Mukherji, D., Junghans, C. & Van der Vegt, N. F. A. Kirkwood–Buff Coarse-Grained Force Fields for Aqueous Solutions. *Journal of Chemical Theory and Computation* **8.** 00013, 1802–1807 (2012).
14. Mullinax, J. W. & Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *The Journal of Chemical Physics* **131,** 104110 (2009).
15. Farah, K., Fogarty, A. C., Böhm, M. C. & Müller-Plathe, F. Temperature dependence of coarse-grained potentials for liquid hexane. *Phys. Chem. Chem. Phys.* **13,** 2894–2902 (2011).
16. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140,** 224104 (2014).
17. Rosenberger, D., Hanke, M. & Vegt, N. F. A. V. d. Comparison of iterative inverse coarse-graining methods. *The European Physical Journal Special Topics* **225,** 1323–1345 (2016).

18. Dunn, N. J. H. & Noid, W. G. Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures. *The Journal of Chemical Physics* **144,** 204124 (2016).

19. Wagner, J. W., Dannenhoffer-Lafage, T., Jin, J. & Voth, G. A. Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions. *The Journal of Chemical Physics* **147,** 044113 (2017).

20. Sanyal, T. & Shell, M. S. Transferable Coarse-Grained Models of Liquid–Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *The Journal of Physical Chemistry B* **122,** 5678–5693 (2018).

21. Jin, J. & Voth, G. A. Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *Journal of Chemical Theory and Computation* **14,** 2180–2197 (2018).

22. Zacharopoulos, N., Vergadou, N. & Theodorou, D. N. Coarse graining using pretabulated potentials: Liquid benzene. *The Journal of Chemical Physics* **122,** 244111 (2005).

23. Mognetti, B. M. *et al.* Coarse-grained models for fluids and their mixtures: Comparison of Monte Carlo studies of their phase behavior with perturbation theory and experiment. *The Journal of Chemical Physics* **130,** 044101 (2009).

24. Brini, E., Herbers, C. R., Deichmann, G. & Van der Vegt, N. F. A. Thermodynamic transferability of coarse-grained potentials for polymer–additive systems. *Physical Chemistry Chemical Physics* **14.** 00006, 11896 (2012).

25. Brini, E. & Van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *The Journal of Chemical Physics* **137.** 00009, 154113 (2012).

26. Murtola, T., Karttunen, M. & Vattulainen, I. Systematic coarse graining from structure using internal states: Application to phospholipid/cholesterol bilayer. *The Journal of Chemical Physics* **131,** 055101 (2009).

27. Rudzinski, J. F. & Noid, W. G. Coarse-graining entropy, forces, and structures. *The Journal of Chemical Physics* **135,** 214101 (2011).

28. Fritz, D., Harmandaris, V. A., Kremer, K. & Van der Vegt, N. F. A. Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules* **42.** 00065, 7579–7588 (2009).

29. *Handbook of fluorous chemistry* (eds Gladysz, J. A., Curran, D. P. & Horváth, I. T.) OCLC: ocm56642600 (Wiley-VCH, Weinheim, 2004).

30. Rühle, V., Junghans, C., Lukyanov, A., Kremer, K. & Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. *Journal of Chemical Theory and Computation* **5,** 3211–3223 (2009).

31. Deichmann, G. & Van der Vegt, N. F. A. Conditional Reversible Work Coarse-Grained Models of Molecular Liquids with Coulomb Electrostatics – A Proof of Concept Study on Weakly Polar Organic Molecules. *Journal of Chemical Theory and Computation* **13,** 6158–6166 (2017).

32. Wagner, J. W., Dama, J. F., Durumeric, A. E. P. & Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *The Journal of Chemical Physics* **145,** 044108 (2016).

33. Qian, H.-J. *et al.* Temperature-Transferable Coarse-Grained Potentials for Ethylbenzene, Polystyrene, and Their Mixtures. *Macromolecules* **41,** 9919–9929 (2008).

34. Binder, K. *et al.* Towards the Quantitative Prediction of the Phase Behavior of Polymer Solutions by Computer Simulation. *Macromolecular Symposia* **278,** 1–9 (2009).

35. Müller, E. A. & Jackson, G. Force-Field Parameters from the SAFT-γ Equation of State for Use in Coarse-Grained Molecular Simulations. *Annual Review of Chemical and Biomolecular Engineering* **5,** 405–427 (2014).

36. Lyubartsev, A., Mirzoev, A., Chen, L. & Laaksonen, A. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss.* **144,** 43–56 (2010).

37. Frenkel, D. & Smit, B. *Understanding molecular simulation: from algorithms to applications* 2nd ed. *Computational science series* **1** (Academic Press, San Diego, 2002).

38. Panagiotopoulos, A., Quirke, N., Stapleton, M. & Tildesley, D. Phase equilibria by simulation in the Gibbs ensemble: Alternative derivation, generalization and application to mixture and membrane equilibria. *Molecular Physics* **63,** 527–545 (1988).

39. Kirkwood, J. G. & Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *The Journal of Chemical Physics* **19,** 774–777 (1951).

40. Krüger, P. *et al.* Kirkwood–Buff Integrals for Finite Volumes. *The Journal of Physical Chemistry Letters* **4,** 235–238 (2013).

41. Milzetti, J., Nayar, D. & Van der Vegt, N. F. A. Convergence of Kirkwood–Buff Integrals of Ideal and Nonideal Aqueous Solutions Using Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **122,** 5515–5526 (2018).

42. Ganguly, P. & Van der Vegt, N. F. A. Convergence of Sampling Kirkwood–Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **9,** 1347–1355 (2013).

43. Galata, A. A., Anogiannakis, S. D. & Theodorou, D. N. Thermodynamic analysis of Lennard-Jones binary mixtures using Kirkwood-Buff theory. *Fluid Phase Equilibria* **470,** 25–37 (2018).

44. Martin, M. G. & Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *The Journal of Physical Chemistry B* **102,** 2569–2577 (1998).

45. Zhang, L. & Siepmann, J. I. Pressure Dependence of the Vapor-Liquid-Liquid Phase Behavior in Ternary Mixtures Consisting of n-Alkanes, n-Perfluoroalkanes, and Carbon Dioxide. *The Journal of Physical Chemistry B* **109,** 2911–2919 (2005).

46. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **17,** 1302–1322 (1978).

47. Deichmann, G. & Van der Vegt, N. F. A. Conditional reversible work coarse-grained models with explicit electrostatics – an application to butylmethylimidazolium ionic liquids. *Physical Chemistry Chemical Physics* **submitted** (2018).

48. Hess, B., Kutzner, C., Van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4,** 435–447 (2008).

49. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81.** 13416, 3684 (1984).

50. Mashayak, S. Y. *et al.* Relative Entropy and Optimization-Driven Coarse-Graining Methods in VOTCA. *PLOS ONE* **10** (ed Huang, X.) e0131754 (2015).

51. Berendsen, H., Van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91,** 43–56 (1995).

52. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2,** 19–25 (2015).

53. Goga, N., Rzepiela, A. J., de Vries, A. H., Marrink, S. J. & Berendsen, H. J. C. Efficient Algorithms for Langevin and DPD Dynamics. *Journal of Chemical Theory and Computation* **8,** 3637–3649 (2012).

54. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52.** 04126, 255–268 (1984).

55. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

56. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).

57. Martin, M. G. MCCCS Towhee: a tool for Monte Carlo molecular simulation. *Molecular Simulation* **39,** 1212–1222 (2013).

58. Martin, M. G. & Siepmann, J. I. Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *The Journal of Physical Chemistry B* **103,** 4508–4517 (1999).

59. Genco, J. M., Teja, A. S. & Kay, W. B. Study of the critical and azeotropic behavior of binary mixtures. 1. Critical states of perfluoromethylcyclohexane-isomeric hexane systems. *Journal of Chemical & Engineering Data* **25,** 350–355 (1980).

60. Mousa, A. E. H. N., Kay, W. B. & Kreglewski, A. The critical constants of binary mixtures of certain perfluoro-compounds with alkanes. *The Journal of Chemical Thermodynamics* **4,** 301–311 (1972).

61. Majer, V. & Svoboda, V. *Enthalpies of vaporization of organic compounds: a crit. review and data compilation Chemical data series* **32.** OCLC: 246396573 (Blackwell Scient. Publ, Oxford, 1985).

62. Lepori, L., Matteoli, E., Spanedda, A., Duce, C. & Tiné, M. R. Volume changes on mixing perfluoroalkanes with alkanes or ethers at 298.15 K. *Fluid Phase Equilibria* **201,** 119–134 (2002).

63. Dallavalle, M. & Van der Vegt, N. F. A. Evaluation of mapping schemes for systematic coarse graining of higher alkanes. *Phys. Chem. Chem. Phys.* **19,** 23034–23042 (2017).

64. Matteoli, E. & Lepori, L. Determination of the excess enthalpy of binary mixtures from the measurements of the heat of solution of the components: application to the perfluorohexane + hexane mixture. *Fluid Phase Equilibria* **174,** 115–131 (2000).

4 Conditional Reversible Work Coarse-Grained Models of Molecular Liquids with Coulomb Electrostatics – a Proof of Concept Study on Weakly Polar Organic Molecules

Abstract

Scale bridging simulations of soft matter rely on the availability of transferable coarse-grained models. In systematic coarse-graining approaches, core principles of statistical mechanics are used to relate the coarse-grained models to the underlying molecular interactions. The conditional reversible work (CRW) method provides effective, nonbonded pair potentials by means of computing coupling free energies between mapped chemical groups. This method has so far been used almost exclusively for systems composed of apolar organic molecules, but additional challenges arise when developing coarse-grained models for polar molecules in which (long-ranged) electrostatic interactions are important. Herein, we present a modified formulation of the CRW method where we divide the effective interaction potential into van-der-Waals and electrostatic components. The shape of the effective electrostatic interaction justifies modeling the electrostatics using a Coulomb potential with point charges on each site that are equal to the net charge of the underlying group of atoms. We perform CRW calculations using two polar molecules as test cases (an ether (1,2-dimethoxyethane) and an ester (ethylpropionate)). The results of subsequent liquid state simulations indicate that the coarse-grained models obtained by the new method are of similar quality with respect to representability and thermodynamic transferability as formerly developed models for apolar systems.

4.1 Introduction

Soft matter research often requires the study of systems and processes on time and length scales that are too large to be modeled by atomistic models. The use of coarse-grained (CG) models enables us to overcome some of these limitations by reducing the number of degrees of freedom in the system and thereby softening the potential energy surface which leads to inherently faster dynamics. [1] A variety of methods have been published in the literature that follow the systematic coarse-graining approach.[2–11] In systematic coarse-graining methods, interactions between sites in the coarse-grained model are based on an underlying fine-grained model in a bottom-up manner. The resulting coarse-grained models have been used for the study of a variety of soft matter systems. [12–16] These methods can be subdivided into two classes: structure based (IBI,[6] IMC[7]) and force-matching (MS-CG,[5] EF-CG,[4] CRW[3]) methods. In structure based coarse-graining, one tries to parametrize the model in such a way that the liquid structure and additional properties are reproduced by the CG model. Force-based methods on the other hand generate a CG potential based directly on the interactions in the fine-grained model. The conditional reversible work (CRW) method developed by Brini et al. is of the latter kind [3], and it is unique because it can be applied by sampling only two molecules in a vacuum environment, thereby providing a CG model at minimal computational cost.

Other CG models have been parametrized in a rather top-down approach with the aim of reproducing e.g. vapor-liquid coexistence curves[17], polymer glass transition properties[18] or bulk partition free energies.[19] The use of CG models is in many cases restricted to state points close to the state point at which they were parametrized. Transferability, i.e. the capability of a model of correctly representing state points beyond the immediate vicinity of the state point of its parametrization, is a desired feature of CG models and several approaches have been published to achieve transferability. These include the use of temperature-dependent potentials[17, 18, 20–22], multistate sampling techniques[23, 24] and other approaches[8, 25, 26].

We have shown in the past that the CRW coarse-graining method provides CG models that are transferable in a rather broad range of temperatures and that these models represent liquid phase structural and thermophysical properties as well as liquid-vapor and liquid-solid interfacial properties in good agreement with experiments and fine-grained simulations.[3, 14, 27] The CRW method has been used to develop CG models for alkanes, polymers and graphene surfaces.[3, 12, 14, 27–29] It has so far not been applied for polar or ionic systems, and in its original formulation it is not well suited for the application to such systems. Herein we present a modified formulation of the CRW method that allows for a separate treatment of electrostatic and Van-der-Waals (vdW) interactions in the CG model. This modified CRW method is subsequently tested on two model systems (an ether and an ester). We would like to point out that it is not the focus of this work to provide CG models whose properties match experimental data. Rather, we would like to show in a "proof-of-principle" manner that the method described herein allows to obtain transferable CG models from underlying fine-grained reference representations of polar molecules.

4.2 Methods

The aim of the systematic coarse-graining procedure is the development of a coarse-grained (CG) model that is based on a fine-grained (FG) reference model (usually an all-atom or united-atom model). The CG model consists of interaction sites that are related to the sites of the FG model by a mapping scheme. In this work, the center of mass mapping is used (see Computational Details).

The main challenge of the coarse-graining process is the determination of suitable non-bonded interaction potentials. In CRW models, interactions between pairs of CG sites are modeled with a free-energy-based pair potential, which corresponds to the reversible work of coupling the microscopic interactions between the atoms contained within the CG interaction sites, given a constant distance between CG sites. This free energy difference can be calculated by the use of a thermodynamic cycle (see fig. 4.1) from the reversible work (RW) associated with the process of moving the two interaction sites from infinity to a finite distance r with direct interactions turned on and off respectively:

$$G_{\text{eff}}(r) = W_{\text{on}}(r) - W_{\text{off}}(r) \tag{4.1}$$

A feature that distinguishes the CRW method from other coarse-graining methods is the fact that the CG potentials can be obtained from atomistic sampling of two molecules in vacuo. Other surroundings can in principle also be chosen[14], but the sampling in vacuo provides a computationally efficient way to calculate the CRW-CG interaction potentials. All CRW potentials used in this work were obtained using vacuum sampling. For a more detailed description of the CRW method we refer the reader to the existing literature[3].

When modeling apolar molecules, the resulting CG potentials will resemble a Lennard-Jones potential with its characteristic features: a steep increase of energy towards zero distance, a potential well, and an attractive tail converging to zero at short distances. The last feature of the potentials is critical for the success of the method as using the aforementioned thermodynamic cycle necessitates that $G_{\text{eff}} \approx 0$ at the largest sampled inter-site distance. For models of polar molecules the interaction potential may contain significant contributions of electrostatic interactions, especially if the mapping scheme is chosen such that interaction sites exhibit net charge $\neq 0$. These electrostatic contributions are typically long-ranged and thus may not have converged to zero at the distance chosen as interaction cutoff. The resulting CRW potential is shifted by an unknown amount, and it is unclear whether it can serve as a physically meaningful interaction potential. Extending the sampling by including more and more distances will, of course, eventually lead to convergence but the larger cutoff length required in the CG simulation will decrease the overall performance. This can be overcome by including a switch function that will force the interactions to zero at a shorter distance value. The choice of this shorter interaction cutoff of this switch function, however, is an additional parametrization choice and may consequently require tuning of the non-bonded interaction potentials.

In this work, we propose a different approach to tackle this challenge: we separate the interaction potential into an effective Van-der-Waals (vdW) and an effective electrostatic (ES) part. This natural decomposition is also used in atomistic force fields and in the effective-force coarse-graining method[4]. With the available algorithms for the treatment of long-range electrostatics, such as particle mesh Ewald and reaction field algorithms, a short cutoff can also be used for the ES part of the interaction potential.

The decomposition of G_{eff} is performed as described in the following. We assert here that non-bonded atomistic interactions in the fine-grained force field underlying our CG model are modeled using a Van-der-Waals and an electrostatic term. We will refer to these terms as the "interaction components" of the fine-grained model. The process of "switching on" these components will be performed in two steps, as opposed to the single step procedure employed in the original CRW method. In the first step, one of the FG interaction components is turned on leading to an intermediate state. The free energy difference associated with this process will be denoted as A_Ψ. Next, the second interaction component is introduced. The free energy difference associated with this change of the FG interactions is denoted B_Φ. In our notation Ψ and Φ serve as placeholders for the interaction components that are being introduced into the model (i.e. ES and vdW). The CG interaction energy is the sum of the two parts:

$$G_{\text{eff}} = A_\Psi + B_\Phi \tag{4.2}$$

The order of introduction can be chosen arbitrarily and the resulting G_{eff} will be independent of this choice. However, one should be aware of the fact that this is not true for its components A and B:

$$A_\Psi \neq B_\Psi \tag{4.3}$$

because the second introduction is performed under the condition that the first interaction component is already present in the system.

The calculation of the free energy differences is almost identical to the original single-step procedure. In addition to W_{on} and W_{off}, we calculate a RW function for the intermediate state in which interaction component Ψ is present in the system ($W_{\text{on},\Psi}$).

The reversible work is calculated by integrating over the projection of the inter-molecular force $\mathbf{F}_{\text{mol}}$ on the unit vector $\mathbf{e}$ parallel to the distance vector connecting the mass centers of the CG sites of interest in constrained molecular dynamics (MD) simulations:

$$W(r) = -\int_\infty^r \langle \mathbf{F}_{\text{mol}} \cdot \mathbf{e} \rangle (r')\, dr' \tag{4.4}$$

To ensure that the distance between the CG interaction sites of interest remains constant throughout the simulation the positions of these sites (α) are constrained to their initial value by modifying the force on all atoms (i) contained within the site in the following way:

$$\mathbf{F}_{i,\alpha} = \mathbf{F'}_{i,\alpha} - \frac{m_i}{m_\alpha} \sum_{j=0}^{N_\alpha} \mathbf{F'}_{j,\alpha} \tag{4.5}$$

where $\mathbf{F'}_{i,\alpha}$ is the (unconstrained) force on atom i, m_i and m_α are the masses of atom i and site α respectively, and N_α is the number of atoms contained in site α. The equation of motion of atoms that are not part of the sites of interest remains unchanged.

These constrained dynamics runs are performed using four different types of interactions between the atoms within the sites of interest and the forces on the molecules are integrated to yield the corresponding reversible work:

1. All interactions present (yielding $W_{\text{on}}(r)$)

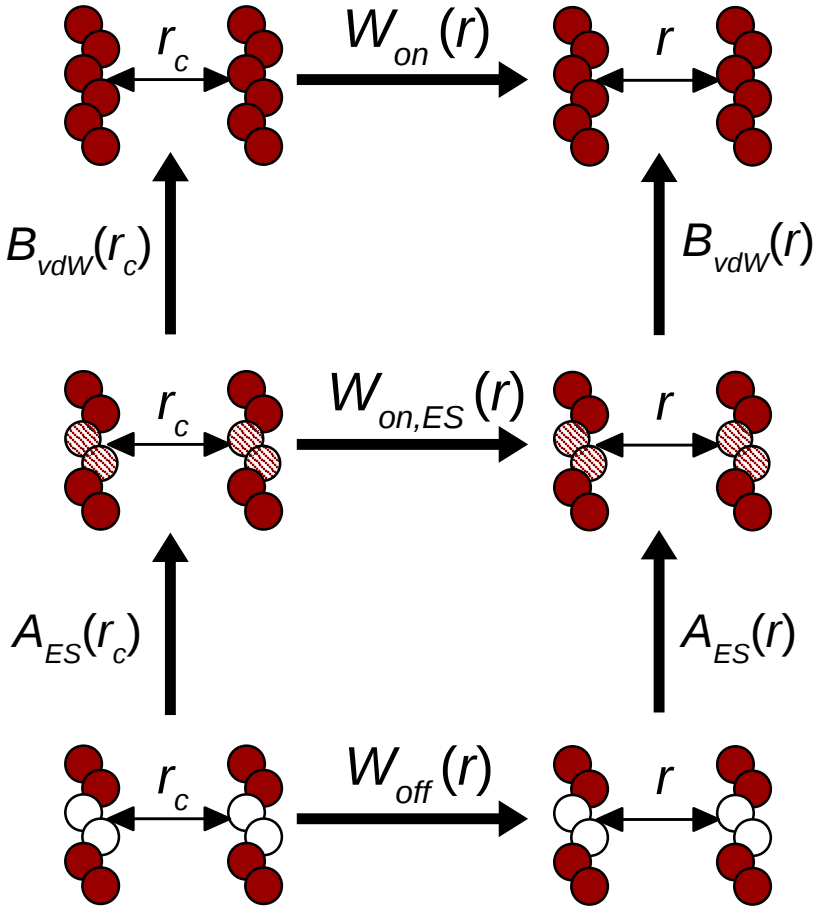


Figure 4.1: Thermodynamic cycles used for the calculation of $A(r)$ and $B(r)$. The two central atoms of a hexamer are merged into one site in the CG model (2:1 mapping). In the lowest part, all interactions between the two central atoms are absent; in the middle part, only the electrostatic component of the interactions is present; and in the upper part, all interactions are included. $A(r)$ and $B(r)$ are then calculated from the potentials of mean force $W(r)$ under the assumption that all $W(r_c) = 0$.

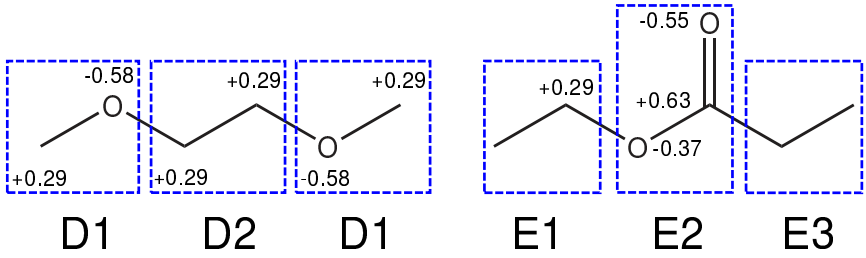


Figure 4.2: Mapping schemes used for the CG models. Partial charges on the united atoms used in the FG simulation are shown to illustrate that there are non-zero net charges on sites D1, D2, E1, and E2.

2. No interactions present (yielding $W_{\mathrm{off}}(r)$)
3. Only electrostatics present (yielding $W_{\mathrm{on,ES}}(r)$)
4. Only Van-der-Waals interactions present (yielding $W_{\mathrm{on,vdW}}(r)$)

From these the CRW interaction potentials are calculated:

$$A_{\Psi} = W_{\mathrm{on},\Psi} - W_{\mathrm{off}} \tag{4.6}$$

$$B_{\Psi} = W_{\mathrm{on}} - W_{\mathrm{on},\Psi} \tag{4.7}$$

For the sake of clarity, we re-emphasize that only those interactions between atoms that are part of the CG sites of interest are modified; all other FG interactions remain unchanged.

This approach provides either $\{A_{\mathrm{ES}}, B_{\mathrm{vdW}}\}$ or $\{A_{\mathrm{vdW}}, B_{\mathrm{ES}}\}$ that can in principle both serve as interaction potentials in our CG simulations. It would be highly advantageous if the electrostatic component can be approximated with a Coulomb potential. Then, established long-range methods for the electrostatics can be used, avoiding a large cutoff radius in the CG simulation. We shall show below that this is indeed possible if we choose A_{ES} as a representation of the electrostatic interactions.

4.3 Computational Details

The two-step CRW method is applied to two molecules, 1,2-dimethoxyethane (DXE) and ethylpropionate (EPP), mapped on a CG representation with three sites as test cases.

Figure 4.2 shows the mapping schemes employed for these molecules. The GROMOS 53A6$_{\mathrm{OXY}}$ force field developed by Horta et al. is used as a fine-grained reference model in the CRW calculations.[30].

4.3.1 Calculation of Non-Bonded CG Interaction Potentials

In order to obtain the interaction potentials for the CG model, calculations are performed using the modified two-step CRW method described above.

Constrained MD runs are performed for all pairs of CG interaction sites using two molecules in vacuo. These simulations were performed using LAMMPS.[31] A time step of 1 fs is used and the total simulation time is 2 ns for each MD run. The temperature is kept at a value of 300 K using the Langevin thermostat[32] with a time constant of 0.5 ps. Distances from 0.3 nm up to a cutoff length of 1.3 nm are sampled with a 0.02 nm distance increment. The electrostatic interactions are modeled by a Coulomb potential and are screened with the experimental relative permittivity as reported in reference 30 (DXE:

7.3, EPP: 5.6) and an infinite cutoff length without long-range corrections. This leads to a closer approximation of the behavior within a liquid phase than the use of unscreened electrostatic interactions would.

It will be demonstrated below that A_{ES} resembles a Coulomb potential. Therefore, rather than A_{ES} itself, an unscreened Coulomb potential (with $\varepsilon_r = 1$) is instead applied in the CG simulations using the net charges of the interaction sites determined by the mapping (fig. 4.2). Long-range electrostatics are modeled using the particle-mesh Ewald (PME) algorithm.[33]

The vdW component in the CG model is modeled by B_{vdW}. In our simulations we do not use a tabulated version of B_{vdW} but instead a Morse potential, described below, fitted to the data. This choice facilitates applications of the reported CG model in which the functional form is used directly, or can be compiled in potential energy tables. A modified Morse potential with 4 parameters has been used to fit B_{vdW}:

$$U(r) = \varepsilon((1 - e^{-k_i(r-r_0)})^2 - 1) \begin{cases} i = 1 & r \leq r_0 \\ i = 2 & r > r_0 \end{cases} \tag{4.8}$$

For very small distances this functional form differs increasingly from the calculated values of B_{vdW}. We therefore limit our fit to distance values for which $B_{\mathrm{vdW}} < 10$ kJ mol^{-1} to obtain a better approximation of the potential energy well and the attractive tail. The error introduced by this procedure is negligible because in an unconstrained simulation the values with considerable deviation from the fitted function are rarely sampled, and the potential will certainly keep its repulsive character when using the functional form. In fig. 4.3, we motivate the choice of this functional form by comparing to a Lennard-Jones function. The Morse potential yields a very good approximation of both the excluded volume and the attractive potential well. The Lennard-Jones functional form does not yield the same degree of accuracy for both of the potentials shown in fig. 4.3. While the attractive region of the potential energy curve can be modeled with the Lennard-Jones functional form, the resulting core repulsion is too steep when compared to the calculated interaction energy. This result is in agreement with findings published by others.[29, 34]

4.3.2 Bonded CG interaction potentials

For DXE the intramolecular interaction potentials are calculated from Boltzmann-inversion of probability distributions gathered from a simulation run of length 100 ns of one molecule in vacuum with all non-bonded interactions excluded. The Langevin thermostat is used with a reference temperature of 300 K and a time constant of 0.5 ps. From the mapped configurations gathered from these simulations we can calculate the potentials for the D1-D1 bond length b and the D1-D2-D1 angle θ:

$$U_B(b) = -k_B T \log \frac{\rho(b)}{b^2} + C_B \tag{4.9}$$

$$U_A(\theta) = -k_B T \log \frac{\rho(\theta)}{\sin(\theta)} + C_A \tag{4.10}$$

where ρ denotes the probability density of the respective intra-molecular degree of freedom and C is a constant chosen such that $\min(U) = 0$. The resulting potentials are shown in figure 4.4.

For EPP the same approach was attempted but as a result of the very rigid geometry of the molecule, the intramolecular potentials obtained from this calculation were too rigid to allow for a stable MD simulation at a timestep of 1 fs. We therefore constrain the bond lengths and angles of EPP to their average value using the LINCS-algorithm[35]. The average bond distances are 0.268 nm and 0.239 nm for the E1-E2 and E2-E3 bonds respectively. The average E1-E2-E3 angle is 124°.

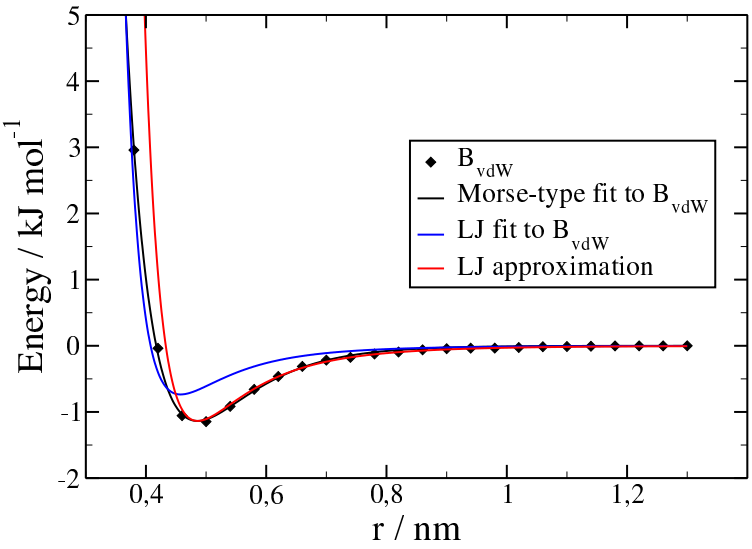


Figure 4.3: Different analytical approximations to B_{vdW} are shown. The modified Morse potential (black) provides the best approximation of the calculated potential. A Lennard-Jones potential obtained by the same fitting procedure (blue curve) does not match the attractive well of the potential. A Lennard-Jones potential that approximates the Morse potential (using $\sigma_{\mathrm{LJ}} = r_0/2^{\frac{1}{6}}$ and $\varepsilon_{\mathrm{LJ}} = \varepsilon$; red curve) reproduces the attractive region but does not match the excluded volume as well as the Morse-type fit.

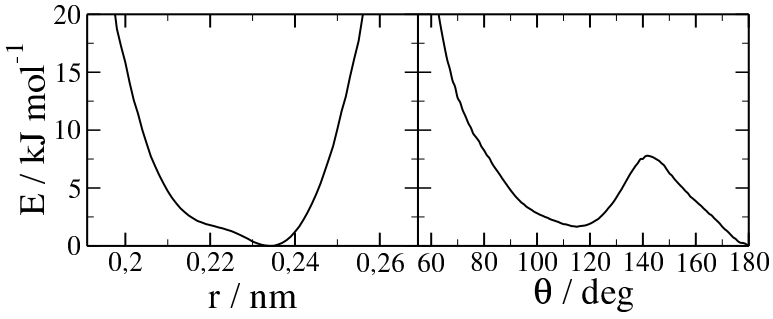


Figure 4.4: Intramolecular potentials used for the DXE model. Left panel: U_B. Right panel: U_A.

4.3.3 Liquid Phase Simulations

Using the FG and CG models, simulations of liquid systems are performed in a cubic simulation box with periodic boundary conditions containing 400 molecules. All liquid phase simulations were performed with a timestep of 1 fs using the GROMACS package version 4.6.7[36]. The simulation boxes are equilibrated for 5 ns using weak-coupling algorithms for temperature and pressure.[37] After equilibration, production runs of 5 ns are performed using the Nosé-Hoover thermostat[38, 39] (time constant 0.2 ps) and Parrinello-Rahman barostat[40, 41] (time constant 2.5 ps; reference pressure 1 bar; compressibility $4.5{\cdot}10^{-5}$ bar^{-1}). The temperature is varied from 260 K to 340 K in increments of 20 K. In order to calculate the surface tension NVT simulations are carried out with a half empty box starting from an equilibrated NPT configuration. The same overall settings are used as for the NPT simulations.

In the FG simulations, we use the twin-range cutoff scheme of the GROMOS force field[42] using a short-range cutoff of 0.8 nm and a long-range cutoff of 1.4 nm. Neighbor lists are updated every 5 steps. A short range cutoff of 1.4 nm is used for the electrostatic interactions. Long-range electrostatic interactions are modeled using the PME algorithm implemented in GROMACS with a grid spacing of 0.12 nm and a relative tolerance of 10^{-5}.[33] No long-range corrections are applied for the Lennard-Jones dispersion contribution to the energy and pressure.

In the CG simulations, a simple cutoff of 1.3 nm is used. The Van-der-Waals interaction is simulated using a potential table generated from equation 4.8 with a spacing of 1 pm and linear interpolation. The short-range cutoff length for the electrostatic interactions is 1.3 nm in the CG model. The long-range part of the ES interactions is treated in the same way as in the FG simulations.

4.4 Results and Discussion

4.4.1 Non-Bonded Interaction Potentials

In figure 4.5, we show a detailed overview of the components of the CRW interaction free energy for the pair D1-D2. This example illustrates the fact that the order of introducing the interaction components is important and why we have chosen the set $\{A_{\mathrm{ES}}, B_{\mathrm{vdW}}\}$ as the final CG force field. Here B_{vdW} is fitted using eq. 4.8 and we can observe a good reproduction of the data by the functional form. The electrostatic component A_{ES} is compared to the Coulomb potential for point charges q_α and q_β shifted to zero at the cutoff distance:

$$U_C(r) = \frac{q_\alpha q_\beta}{4\pi\varepsilon_0\varepsilon_r}\left(\frac{1}{r} - \frac{1}{r_c}\right) \tag{4.11}$$

The values of q_α and q_β are set equal to the net charge on the sites here and in all following plots and are defined as:

$$q_\alpha = \sum_{j=0}^{N_\alpha} q_{j,\alpha} \tag{4.12}$$

where $q_{j,\alpha}$ is the partial charge on atom j in CG site α and N_α is the number of atoms contained in CG site α.

Only a good match of the calculated CRW potential to this functional form justifies the approximation to model A_{ES} with a simple Coulomb potential. The smaller the inter-site distance, the more deviation from this behavior can be expected. Indeed, fig. 4.5 shows that an increasing difference between the theoretical form and the CRW potential is observed. However, the overall agreement of the Coulomb potential and the CRW data suggests that the point-charge approximation of the CG sites can be justified. The energy components of the set $\{A_{\mathrm{vdW}}, B_{\mathrm{ES}}\}$ do not have these characteristic physical shapes. The electrostatic component is repulsive and the convergence of the attractive tail of the Van-der-Waals part does

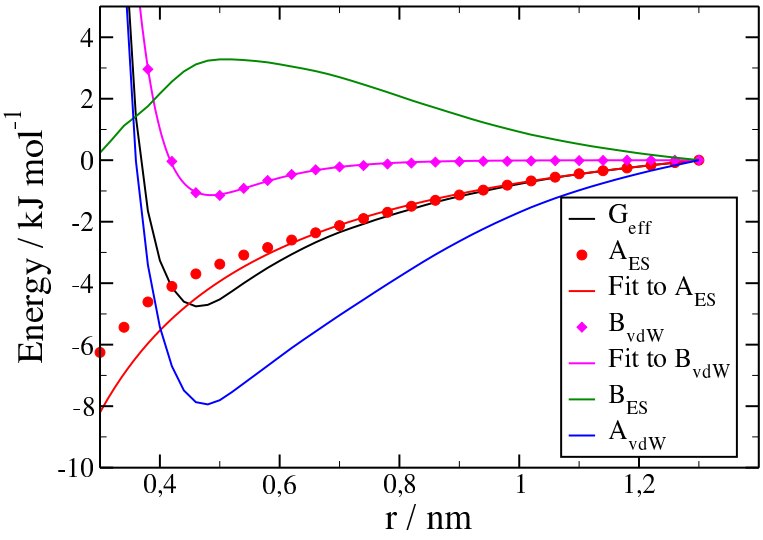


Figure 4.5: A detailed illustration of the interaction energies obtained using both possible introduction orders for the pair of CG sites D1 and D2. A_{ES} can be approximated by a Coulomb potential and B_{vdW} is fitted using a Morse-type potential.

not converge to zero on short distances. In this case, the interpretation of the interaction components is not as straightforward and we could not justify the use of point charges from the physical form of the interaction potential.

Figures 4.6 and 4.7 show the CG interaction potentials for the remaining pairs of DXE and EPP. Here we show the fitted Van-der-Waals component and the calculated CRW potential for the electrostatic component (symbols) as well as for the theoretical Coulomb interaction potential based on point charges (lines). Table 4.1 gives an overview of all the non-bonded parameters used in the CG force field.

4.4.2 Physical Properties of the Coarse-Grained Model Systems

In this section, we present results from liquid phase simulations using the CRW-CG models described above and compare them to quantities obtained from simulations using the FG reference force field.

The properties investigated include the isobaric heat capacity c_p and the isothermal compressibility κ, which can be calculated from enthalpy and volume fluctuations, respectively, in the constant pressure-temperature ensemble:

$$c_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{k_\mathrm{B} T^2 N_\mathrm{DOF}} \tag{4.13}$$

$$\kappa = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_\mathrm{B} T \langle V \rangle} \tag{4.14}$$

where V denotes the total volume, T the temperature and k_B the Boltzmann constant. The total enthalpy $H = E_\mathrm{kin} + E_\mathrm{pot} + pV$ is calculated using the pressure p and the total kinetic (E_kin) and potential (E_pot) energies of the system. The number of thermal degrees of freedom $N_\mathrm{DOF} = 3N_\mathrm{at} - 3$, where N_at denotes the total number of atoms in the simulation.

In addition, we calculate the relative dielectric permittivity ε_r of the liquid using[43]:

$$\varepsilon_r = 1 + \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{3\varepsilon_0 V k_\mathrm{B} T}, \tag{4.15}$$

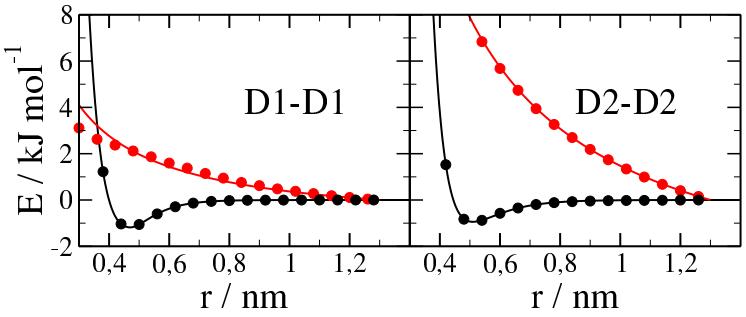


Figure 4.6: CG interaction potentials of the like pairs of DXE. The black circles represent the calculated B_{vdW} potential, while the black curve shows the fitted B_{vdW} potential. Red circles represent the calculated A_{ES} and the red curves show the theoretical prediction for a potential of point charges.

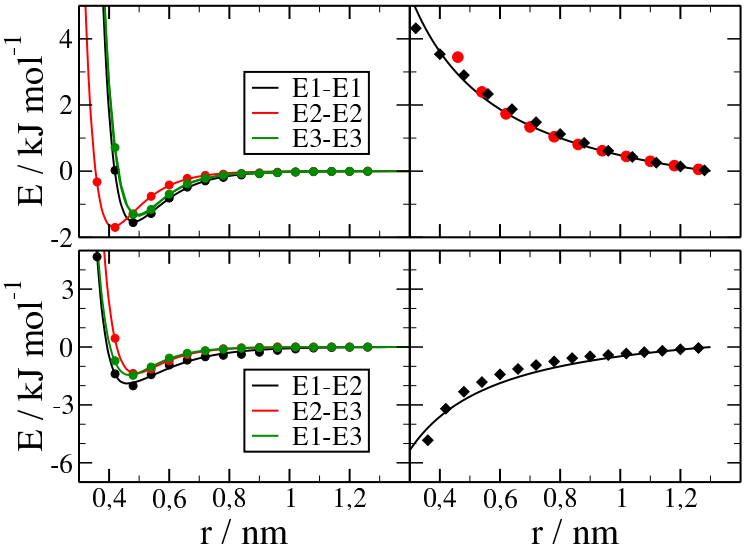


Figure 4.7: CG interaction potentials for all pairs in the EPP CG model. The curves in the left half show the fitted functions of B_{vdW}. Circles represent the calculated values of B_{vdW}. In the right half circles represent the calculated A_{ES} and lines show the theoretical prediction. Like pairs are shown in the upper half, unlike pairs in the lower half respectively. Note that the theoretically predicted potential is identical for the two like pairs and that for all pairs containing E3 $A_{\mathrm{ES}} = 0$ (not shown here).

Table 4.1: Non-bonded parameters of the CG force field

	ε / kJ mol^{-1}	r_0 / nm	k_1 / nm^{-1}	k_2 / nm^{-1}	q / e
DXE					
D1-D1	1.182	0.471	9.739	13.583	-0.29
D2-D2	0.942	0.510	10.731	10.433	0.58
D1-D2	1.136	0.486	10.065	10.659	
EPP					
E1-E1	1.529	0.481	11.136	9.537	0.29
E2-E2	1.678	0.413	11.934	10.538	-0.29
E3-E3	1.315	0.495	10.550	10.841	0
E1-E2	1.887	0.458	10.828	7.348	
E2-E3	1.451	0.469	10.592	11.014	
E1-E3	1.362	0.489	10.867	10.920	

Table 4.2: Properties at 300K

	DXE		EPP	
Property	FG	CG	FG	CG
ρ / kg m^{-3}	871	919	855	930
γ / mN m^{-1}	29.9 ± 0.4	26.2 ± 1.1	22.2 ± 0.4	26.8 ± 1.7
κ / 10^{-9} Pa^{-1}	0.97 ± 0.01	1.17 ± 0.02	1.18 ± 0.01	1.03 ± 0.04
c_p / J mol^{-1} K^{-1}	29.7 ± 0.1	16.7 ± 0.3	29.2 ± 0.1	11.5 ± 0.7
ε_r	10.44	30	4.12	10.7
G_k	1.15	2.8	1.05	1.25

where $\mathbf{M}$ is the total dipole moment of the system and ε_0 is the permittivity of the vacuum.

Another quantity relating to the electrostatic properties of the liquid is the finite system Kirkwood factor G_k which relates to the strength of correlation between molecular dipole moments and can be calculated as follows:[44]

$$G_k = \frac{\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2}{N_{\mathrm{mol}} \langle \mu^2 \rangle} \tag{4.16}$$

using the number of molecules N_{mol} and the molecular dipole moment μ.

From NVT simulations with a half empty box (extended to length L_z in z-direction) we calculate the surface tension from the pressure tensor components p_{ii}:

$$\gamma = \frac{L_z}{2} \left(p_{zz} - \frac{p_{xx} + p_{yy}}{2} \right) \tag{4.17}$$

Dimethoxyethane

The RDF of all interaction pairs are shown in figure 4.8. While the RDF of the D1-D1 pair is almost quantitatively reproduced, the CG model yields only a rough qualitative reproduction for the remaining two pairs. Especially for the D2-D2 pair we can observe a loss of resolution and an increase in long-range correlations which is probably due to an increase in the effective excluded volume of the central methylene group. The CG model reproduces the excluded volume of the D1-D2 pair well but also fails

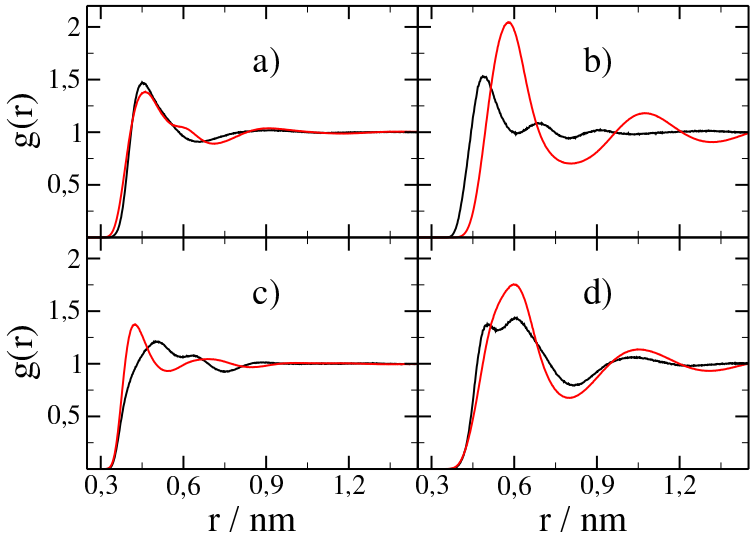


Figure 4.8: Radial distribution functions of the centers of mass of CG sites from simulations using the FG (black) and CG (red) models. a) D1-D1; b) D2-D2; c) D1-D2; d) RDF of molecular COM.

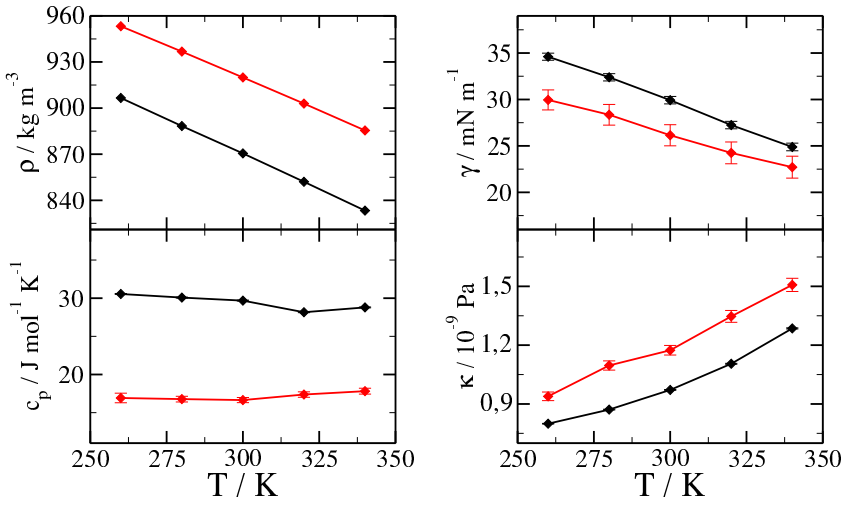


Figure 4.9: Calculated physical properties of liquid systems of DXE modeled with the FG (black) and CG (red) models as a function of temperature.

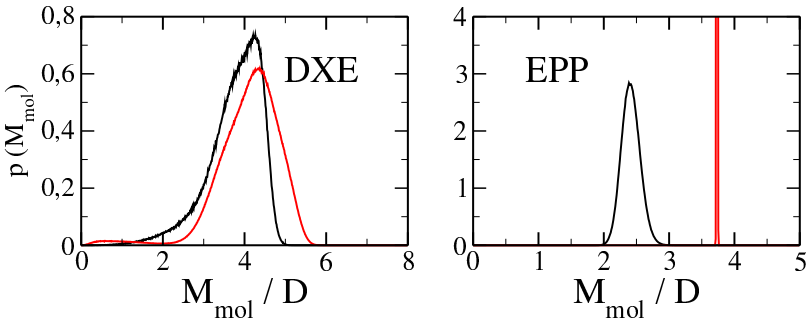


Figure 4.10: Distribution of molecular dipole moments in the FG (black) and CG (red) simulations of DXE and EPP.

to resolve details of the RDF of this pair qualitatively and quantitatively. The same can also be observed in the RDF of the molar centers of mass.

The values of the thermodynamic observables computed for the FG and CG models are shown in table 4.2 for a temperature of 300 K and in figure 4.9 for a range of temperatures.

The mass density of the liquid is overestimated by 5% at 300 K and the thermal expansion coefficient of the FG model is well reproduced by the CG model in the temperature range between 260 K and 340 K.

The heat capacity per degree of freedom predicted by the CG model is lower than the value calculated from FG reference simulations. This result is not surprising because the coarse-graining procedure softens the potential energy landscape of the system and therefore will reduce its capability to store thermal energy in its translational and internal degrees of freedom.

The isothermal compressibility is overestimated at all temperatures but the temperature dependence is reproduced by the CG model. We attribute this overestimation to an increase in volume fluctuations which are of higher magnitude in the CG model due to a softened potential energy surface. We would like to point out that a more accurate representation of the volume fluctuations of the system is in principle possible by the use of the barostat formulation for CG systems published recently by Dunn and Noid. [45] The use of this thermostat, however, requires additional parametrization that would go beyond the scope of this work.

The liquid-vapor surface tension for DXE is underestimated by $\sim$ 10% and the temperature dependence is well reproduced in the given temperature range. These results document the excellent state point transferability of the CRW-CG models as reproduction of surface properties requires a very good description of both liquid and vapor behavior.

The value of ε_r is overestimated in the CG model although the molecular dipole moment distribution (shown in fig. 4.10, left panel) of the FG model is well reproduced by the CG model. A comparison of the Kirkwood factors shows that G_k is increased in the CG model in comparison to the FG reference by approximately the same ratio as ε_r. This indicates that dipole-dipole correlations are too strong which may correlate with the overestimation of D2-D2 density correlations shown in fig. 4.8b.

Ethylpropionate

The $g(r)$ of all interaction pairs are shown in figure 4.11. The performance of the EPP CG model in reproducing the liquid structure is better than that of the DXE model. We observe a match of the RDF for all pairs and although a perfect quantitative reproduction is not achieved (and should not

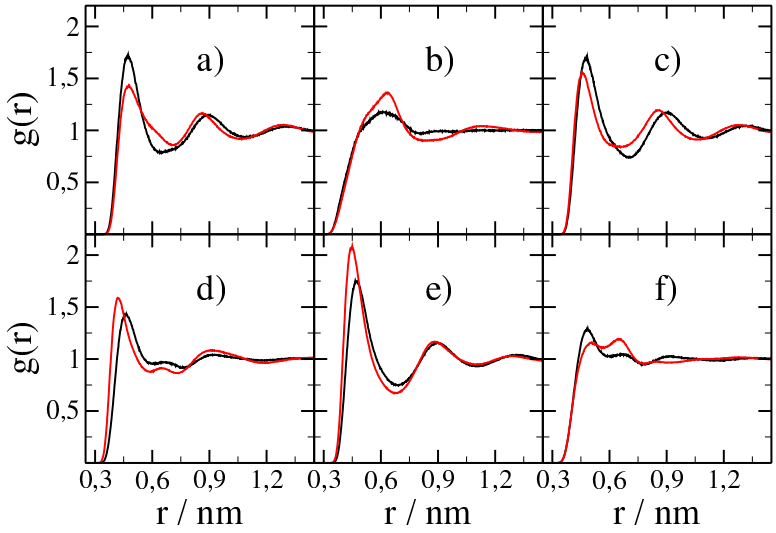


Figure 4.11: Radial distribution functions of the centers of mass of CG sites from simulations using the FG (black) and CG (red) models. a) E1-E1; b) E2-E2; c) E3-E3; d) E1-E2; e) E1-E3; f) E2-E3.

be expected anyway) the CG model is able to reproduce all essential features of the reference liquid structure qualitatively.

The mass density is slightly higher compared to the FG reference (overestimation of $\sim$ 9% at 300K) and the thermal expansion coefficient is less well matched by the CG model which predicts a lower expansion coefficient than the reference FG simulation.

The heat capacity is lower in the CG model, a finding that has already been discussed above for DXE.

The isothermal compressibility is underestimated for all temperatures and the temperature dependence is also different for the compared models. This finding is not intuitive because it might be expected that the softened potential energy surface in the CG model leads to increased volume fluctuations like it was observed for DXE. However we suspect the reason for the underestimation of κ is to be found in the higher density in the CG model which renders the liquid less compressible. The density is overestimated more strongly for EPP than for DXE and therefore density effects might trump the effect of the softer interactions. This hypothesis is supported by the fact that the mismatch in κ is larger at temperatures that show a larger mismatch in the liquid density.

The liquid-vapor surface tension prediction by the two investigated models differ by $\sim$ 15% and the temperature dependence is well reproduced.

The relative permittivity ε_r is overestimated in the CG simulations by a factor of 2.5 although the Kirkwood factor G_k is reproduced better than in the case of DXE. However we observe that the molecular dipole moment is overestimated in the CG model (fig. 4.10). This suggests that the reason for the mismatch in the relative permittivity is not so much due to increased orientational correlations between dipoles but rather due to the increase of the molecular dipole moment itself.

4.5 Conclusion

We have presented a modification to the CRW bottom-up coarse-graining method that allows for a separation of electrostatic and Van-der-Waals components in the CG model by stepwise introduction of the respective fine-grained interaction components. This procedure has been applied using two polar organic molecules as test cases. The effective pair potentials obtained can be decomposed into effective

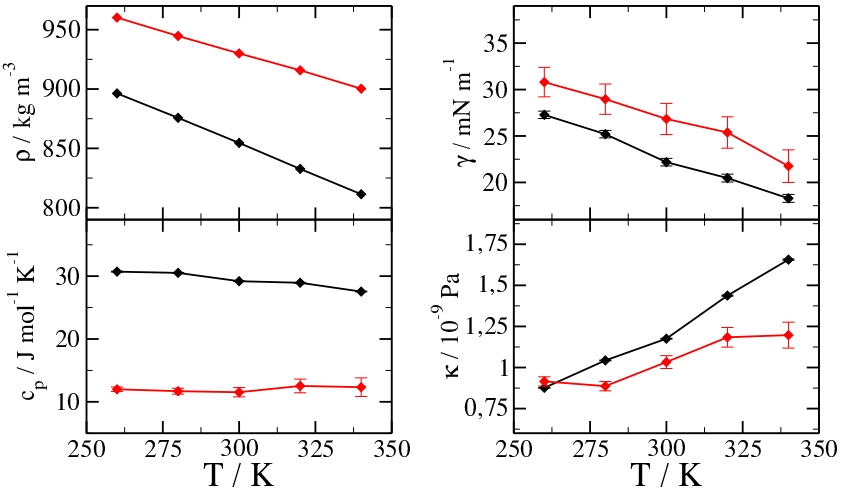


Figure 4.12: Calculated physical properties of liquid systems of EPP modeled with the FG (black) and CG (red) models as a function of temperature.

electrostatic and effective vdW interactions only when choosing the appropriate order of introducing the fine-grained interaction components. In simulations of the CG system, the new approach ascertains that the vdW part of the interaction free energy has converged at the cutoff. Furthermore, methods for long-range electrostatics can be used for a proper treatment of the electrostatic interactions of the model. The CRW-CG interaction potentials are derived from two fine-grained molecules in vacuo and are not parametrized to fit any structural or thermodynamic target quantity. Still, coarse-grained simulations reproduce liquid-state properties in satisfactory agreement with the fine-grained model. In particular, the bulk density and predict the thermal expansion coefficient are reproduced quite accurately. This shows that CRW models with explicit electrostatics show a high degree of transferability, as has been observed for CRW models of apolar molecules.[3, 12, 14] Furthermore, both CG models presented in this work predict the vapor-liquid surface tension in good agreement with reference simulations, an observation that can serve as further evidence for the state point transferability of CRW-CG potentials. Unfortunately, the CRW models do not reproduce the relative dielectric permittivity of the FG systems. This quantity depends on a number of factors and it is unclear whether modifying one of these factors independently will result an a better reproduction of the dielectric properties. However, if a correct reproduction of this quantity is desired one could attempt to tune the charges on the CG sites in order to change the molecular dipole moment in systems where the orientational correlations between dipole moments are well reproduced (like EPP). However, we advice to perform such tuning carefully because whenever such a modification of the model is performed, the interaction free energy G_{eff} will no longer be independent of the introduction order and it is likely that the model will lose a great deal of its predictive power and transferability when such *a posteriori* modifications are applied.

We conclude that the modified CRW method presented in this work is a good extension to the toolbox of systematic bottom-up coarse-graining of polar molecular systems. The method can be especially useful for polymers or systems with a large number of different interaction sites. For these systems vacuum sampling is particularly advantageous as liquid state simulations will be computationally expensive and structure-based CG methods may be difficult to converge. Further studies remain to be performed to test the method on more polar molecules (e.g. alcohols, amines, ionic liquids) in order to determine, whether CG models can be obtained by our modified method also for these systems.

We finally point out that the effective-force coarse-graining method introduced by Wang and coworkers uses an approach that is methodologically very similar to the CRW method.[4] In their study, the authors also show that the CG interactions can be decomposed into effective Van-der-Waals and electrostatic contributions. As parametrization of EF-CG methods can be efficiently performed in the liquid phase differences can be expected with respect to the CRW potentials originating from vacuum sampling. Another methodological difference, however, lies within the fact that effective potentials derived with the CRW method contain indirect contributions originating from the modified sampling in a simulation with excluded interactions. In EF-CG, the sampling is only performed with all interactions present in the system and therefore these contributions are absent. As a result of this subtle methodological difference we can expect slightly different potentials even if the fine-grained sampling is performed in an identical environment. The resulting CG potentials from the EF-CG and CRW methods will however show a similar shape, and it would be interesting to see to which degree the resulting models are comparable in terms of transferability and representability.

4.6 References

1. Peter, C. & Kremer, K. Multiscale simulation of soft matter systems. *Faraday Discussions* **144.** 00053, 9 (2010).
2. Brini, E. *et al.* Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter* **9.** 00021, 2108 (2013).
3. Brini, E., Marcon, V. & Van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics* **13.** 00022, 10468 (2011).
4. Wang, Y., Noid, W. G., Liu, P. & Voth, G. A. Effective force coarse-graining. *Physical Chemistry Chemical Physics* **11,** 2002 (2009).
5. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128,** 244114 (2008).
6. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24.** 00428, 1624–1636 (2003).
7. Lyubartsev, A. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52.** 00390, 3730–3737 (1995).
8. Ganguly, P., Mukherji, D., Junghans, C. & Van der Vegt, N. F. A. Kirkwood–Buff Coarse-Grained Force Fields for Aqueous Solutions. *Journal of Chemical Theory and Computation* **8.** 00013, 1802–1807 (2012).
9. Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129,** 144108 (2008).
10. Mullinax, J. W. & Noid, W. G. A Generalized-Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials †. *The Journal of Physical Chemistry C* **114,** 5661–5674 (2010).
11. McCarty, J., Clark, A. J., Copperman, J. & Guenza, M. G. An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *The Journal of Chemical Physics* **140,** 204913 (2014).
12. Brini, E., Herbers, C. R., Deichmann, G. & Van der Vegt, N. F. A. Thermodynamic transferability of coarse-grained potentials for polymer–additive systems. *Physical Chemistry Chemical Physics* **14.** 00006, 11896 (2012).
13. Fritz, D., Herbers, C. R., Kremer, K. & Van der Vegt, N. F. A. Hierarchical modeling of polymer permeation. *Soft Matter* **5,** 4556 (2009).

14. Brini, E. & Van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *The Journal of Chemical Physics* **137.** 00009, 154113 (2012).

15. Wang, Y., Feng, S. & Voth, G. A. Transferable Coarse-Grained Models for Ionic Liquids. *J. Chem. Theory Comput.* **5,** 1091–1098 (2009).

16. Saunders, M. G. & Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **42,** 73–93 (2013).

17. Cao, F. & Sun, H. Transferability and Nonbond Functional Form of Coarse Grained Force Field – Tested on Linear Alkanes. *Journal of Chemical Theory and Computation* **11,** 4760–4769 (2015).

18. Hsu, D. D., Xia, W., Arturo, S. G. & Keten, S. Systematic Method for Thermomechanically Consistent Coarse-Graining: A Universal Model for Methacrylate-Based Polymers. *J. Chem. Theory Comput.* **10,** 2514–2527 (2014).

19. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).

20. Qian, H.-J. *et al.* Temperature-Transferable Coarse-Grained Potentials for Ethylbenzene, Polystyrene, and Their Mixtures. *Macromolecules* **41,** 9919–9929 (2008).

21. Mognetti, B. M. *et al.* Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide. *J. Chem. Phys.* **128,** 104501 (2008).

22. Mognetti, B. M. *et al.* Coarse-grained models for fluids and their mixtures: Comparison of Monte Carlo studies of their phase behavior with perturbation theory and experiment. *The Journal of Chemical Physics* **130,** 044101 (2009).

23. Mullinax, J. W. & Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *The Journal of Chemical Physics* **131,** 104110 (2009).

24. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140,** 224104 (2014).

25. Hess, B., Holm, C. & Van der Vegt, N. Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Chem. Phys.* **124,** 164509 (2006).

26. Villa, A., Peter, C. & Van der Vegt, N. F. A. Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water. *J. Chem. Theory Comput.* **6,** 2434–2444 (2010).

27. Ardham, V. R., Deichmann, G., Van der Vegt, N. F. A. & Leroy, F. Solid-liquid work of adhesion of coarse-grained models of n-hexane on graphene layers derived from the conditional reversible work method. *J. Chem. Phys.* **143,** 243135 (2015).

28. Fritz, D., Harmandaris, V. A., Kremer, K. & Van der Vegt, N. F. A. Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules* **42.** 00065, 7579–7588 (2009).

29. Dallavalle, M. & Van der Vegt, N. F. A. Evaluation of mapping schemes for systematic coarse graining of higher alkanes. *Phys. Chem. Chem. Phys.* **19,** 23034–23042 (2017).

30. Horta, B. A. C., Fuchs, P. F. J., Van Gunsteren, W. F. & Hünenberger, P. H. New Interaction Parameters for Oxygen Compounds in the GROMOS Force Field: Improved Pure-Liquid and Solvation Properties for Alcohols, Ethers, Aldehydes, Ketones, Carboxylic Acids, and Esters. *J. Chem. Theory Comput.* **7,** 1016–1031 (2011).

31. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117.** 05802, 1–19 (1995).

32. Schneider, T. & Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **17,** 1302–1322 (1978).

33. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

34. Hills, R. D., Lu, L. & Voth, G. A. Multiscale Coarse-Graining of the Protein Energy Landscape. *PLoS Comput. Biol.* **6** (ed Nussinov, R.) e1000827 (2010).

35. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).

36. Hess, B., Kutzner, C., Van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4,** 435–447 (2008).

37. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81.** 13416, 3684 (1984).

38. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52.** 04126, 255–268 (1984).

39. Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (1985).

40. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

41. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50,** 1055–1076 (1983).

42. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **25,** 1656–1676 (2004).

43. Neumann, M. Computer simulation and the dielectric constant at finite wavelength. *Mol. Phys.* **57,** 97–121 (1986).

44. Glättli, A., Daura, X. & Van Gunsteren, W. F. Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *J. Chem. Phys.* **116,** 9811–9828 (2002).

45. Dunn, N. J. H. & Noid, W. G. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *The Journal of Chemical Physics* **143,** 243148 (2015).

5 Conditional Reversible Work Coarse-Grained Models with Explicit Electrostatics - an Application to Butylmethylimidazolium Ionic Liquids

Abstract

Parameterization of nonbonded interactions is fundamental in the development of new coarse-grained (CG) models. The conditional reversible work (CRW) method is unique among the systematic coarse-graining methods since it requires only the simulation of two molecules in vacuo, making parametrizations of new models computationally inexpensive. This method has been applied successfully to apolar systems and has recently been extended for the parametrization of models with explicit electrostatics. In this work, the ECRW method is used to parameterize models for butylmethylimidazolium (BMIM) ionic liquids with three different anions. Results of subsequent CG simulations are used to illustrate the strengths and weaknesses of the resulting models. In particular, the structural and dynamical properties of the bulk, as well as the behavior at the liquid-vapor interface, are studied. It is shown that the ECRW coarse-graining method produces CG models of high quality for all selected test systems.

5.1 Introduction

Coarse-grained (CG) models are of great importance in molecular simulations.[1, 2] Using molecular models which are not based on an atomistic description but use a lower resolution increases the computational performance of a simulation. Apart from this obvious advantage of coarse-graining, we can learn, by evaluating the quality of models based on different representations of the same system, about the degrees of freedom required to realistically model thermodynamic, structural, and dynamic quantities.

In systematic coarse-graining methods, a model is generated based on a finer (usually atomistic) representation of the system.[2] The potential energy surface of a coarse-grained system is characterized by the multi-body potential of mean force (PMF). The aim of bottom-up CG modeling is to approximate the multi-body PMF by a set of pairwise additive effective potentials. For the calculation of these effective potentials a variety of methods have been published that are based either on structural correlations or on forces in the fine-grained (FG) system.[2–12]

The quality of a CG model can be judged using two aspects: representability and transferability. Representability describes the quality of a model in reproducing the system at a state point at which it was parameterized[13–15], while transferability describes the capability of a model to correctly reproduce the behavior of the system at different state points. Transferability can be introduced into a model by a variety of *a posteriori* fits and additional model parameters (e.g. temperature-dependent interaction potentials)[9, 16–24]. Other approaches have focused more on the possibility to design the coarse-graining procedure such that the transferability of the model emerges as a natural consequence of the manner in which the interaction potentials are calculated.[3–5, 25–34]

Systematic attempts to achieve state point transferability have been undertaken especially on the basis of the multiscale coarse-graining (MS-CG) method[35, 36] which provides an approach to approximate the multi-body potential of mean force through pair potentials. It has been shown that at a constant state point, the MS-CG method provides an accurate and well-informed approximation to the multi-body PMF.[6] Thermodynamic transferability of CG models parameterized through the MS-CG method has been achieved by the use of temperature rescaling[25] and extended ensemble sampling approaches.[22, 28, 37] Transferability with respect to density has been achieved by the use of density dependent CG potentials.[34, 38, 39] The behavior near an interface can be described by models using order parameter dependent CG potentials.[40]

Another coarse-graining paradigm, the method of ultra coarse-graining (UCG), has been introduced with the aim of parametrizing coarse-grained yet transferable models by combining the advantages of systematic bottom-up coarse-graining with the state-based approach behind Markov state models. [31, 41, 42] A UCG model contains information on substates of the CG sites gathered from an analysis of the FG reference system. It has been shown recently, that this coarse-graining method also can be applied to model the behavior at interfaces correctly.[32]

In this work, we will use the conditional reversible work (CRW) method for coarse-graining.[3] Here, the forces between two molecules in vacuo are used to exclude interfering multi-body correlations originating from the surrounding molecules, which are assumed to be among the leading causes of state-point dependence of CG models. Models derived by the CRW method have been shown to reproduce structure and thermodynamics of various systems in a transferable manner.[3, 4, 43–46]

While the original CRW method was developed for the parametrization of apolar systems, we recently published a modified version of the method. In this extended approach (which we shall refer to as ECRW in the following), we obtain separate effective interaction potentials for the Van-der-Waals and electrostatic components in the model.[4] We have shown in a model study that this approach leads to a transferable CG force field for weakly polar molecular liquids. Because multi-body effects are expected to have a large impact on the properties of polar fluides, it remains to be explored if the pairwise interaction Ansatz of the ECRW approach provides CG models of equally good quality for highly polar systems. The purpose of this article is to demonstrate that the ECRW method is also suitable for the parametrization of force fields for components with a strong polar character. To this end, we parameterize force fields for three different butylmethylimidazolium ionic liquids (BMI-IL).

Room temperature ionic liquids (RTIL) have been of increasing interest in experimental research, as well as in molecular simulations.[47–52] Because of their high viscosity at room temperature, coarse-grained models are of great interest for the simulations of these systems. The long time scales required to equilibrate dense and viscous systems, such as RTIL, are more efficiently equilibrated using coarse-grained models.

The challenge of modeling RTIL in simulations lies in the correct description of the electrostatics; several methods and models have been brought forward addressing this challenge.[53] Since there are multiple coarse-grained models for BMI-IL in the literature,[54–58] the intent of this work is not merely to add to this variety. Rather we would like to point out a way to easily and efficiently parameterize CG models for ionic liquids or similar systems that can be applied in future applications, where a CG model for the system at hand may not be available. Since the ECRW method is cheap, it may be used to inexpensively compare mapping schemes and to gain insight into the interdependence of coarse-grained representation and model quality for a given system.

The transferability of the CG models is evaluated herein with respect to two factors. First, we compare the mass density at different temperatures to investigate the temperature transferability. Second, we calculate properties of the liquid-vapor interface. A good reproduction of the behavior at interfaces is another illustration of transferability since the interactions near the interface are substantially different to those in the bulk. Coarse-grained modeling of interfacial properties is a challenging task[32, 40, 59–61] and CRW models have been shown to yield good results for the simulation of systems at interfaces.[4, 44, 46]

The remainder of this paper is structured as follows: In section 5.2, we shortly reiterate the essential steps of the ECRW method and introduce a novel technique of generating the trajectories required for the model parametrization. In section 5.3, the resulting models are presented. In sections 5.4 and 5.5, the parameters and results of subsequent simulations in the CG and FG representation are presented and discussed. Section 5.6 contains concluding remarks.

5.2 Methods

5.2.1 Extended Conditional Reversible Work Coarse-Graining

We can obtain a coarse-grained model in a bottom-up manner using a fine-grained (i.e., atomistic) reference model and a mapping scheme that relates the fine-grained coordinates to the coarse-grained ones. In this work the mapping scheme is a center-of-mass mapping that defines the location of a CG site at the center of mass of topologically close atoms. Each fine-grained atom is part of exactly one CG site.

The parametrization of the CG model is performed using the extended (two-step) conditional reversible work (ECRW) method for polar systems. In the following, a brief summary of the method will be given. For a detailed description we refer the reader to the existing literature.[4] In ECRW, the pair interaction between CG sites i and j is modeled using effective Van-der-Waals and effective electrostatic interactions ($A_{\mathrm{vdW}}(r_{ij})$ and $B_{\mathrm{ES}}(r_{ij})$). These interaction components are coupling free energies that correspond to the process of "switching-on" the respective interaction component for the atoms contained in the respective CG sites of interest, while these sites are fixed in space and the distance r_{ij} between them is constant. The interaction potential is calculated by means of a thermodynamic cycle (fig. 5.1) using the reversible work W done against the intermolecular forces when two molecules approach from the cut-off distance to the finite distance r_{ij}. We obtain W by integrating the constraint force F_{C}, i.e., the projection of the intermolecular force on the i-j distance vector:

$$F_{\mathrm{C}}(r_{ij}) = \left\langle \sum_{a\in\mathrm{Mol.1}} \sum_{b\in\mathrm{Mol.2}} \frac{\mathbf{F}_{ab}\cdot\mathbf{r}_{ij}}{|\mathbf{r}_{ij}|} \right\rangle \tag{5.1}$$

$$W(r_{ij}) = -\int_{r_{\mathrm{c}}}^{r_{ij}} F_{\mathrm{C}}(r')dr', \tag{5.2}$$

where $\mathbf{F}_{ab}$ is the pair force between atoms a and b. W is sampled in different states which we denote with the following subscripts: W_{On} (all pair interactions present in the model), $W_{\mathrm{On,vdW}}$ (only vdW interactions are present between atoms contained in the CG sites of interest), and W_{Off} (no interactions between atoms contained in the CG sites of interest). The effective interaction potentials are obtained from W sampled in these different states (compare fig. 5.1 for a schematic depiction of the process):

$$A_{\mathrm{vdW}} = W_{\mathrm{On,vdW}} - W_{\mathrm{Off}} \tag{5.3}$$

$$B_{\mathrm{ES}} = W_{\mathrm{On}} - W_{\mathrm{On,vdW}} \tag{5.4}$$

The indices in the lowest two equations imply a procedure of turning on the Van-der-Waals component first and the electrostatic component second (see fig. 5.1).

The energy components A_{vdW} and B_{ES} calculated in this manner are fitted using the functional forms introduced in reference [4]. For A_{vdW} the potential is approximated using a Morse-type potential with two stiffness parameters, k_1 and k_2, controlling the stiffness of the repulsive and attractive parts, respectively:

$$U_{\mathrm{vdW}}(r) = \varepsilon((1-e^{-k_i(r-r_0)})^2 - 1)\begin{cases} i=1 & \text{if } r \le r_0 \\ i=2 & \text{if } r > r_0 \end{cases}, \tag{5.5}$$

where ε is the potential well depth and r_0 is the minimum energy distance. The choice of this functional form has been discussed in our earlier article on the ECRW method.[4] The two-term Morse potential combines a certain simplicity (as it has only four parameters which with a somewhat intuitive meaning) with the flexibility required to fit the potential curves emerging from CRW calculations. The main benefit of this functional form in the context of this work is the possibility to modify the repulsive part of the potential without changing the attractive part (and vice versa).

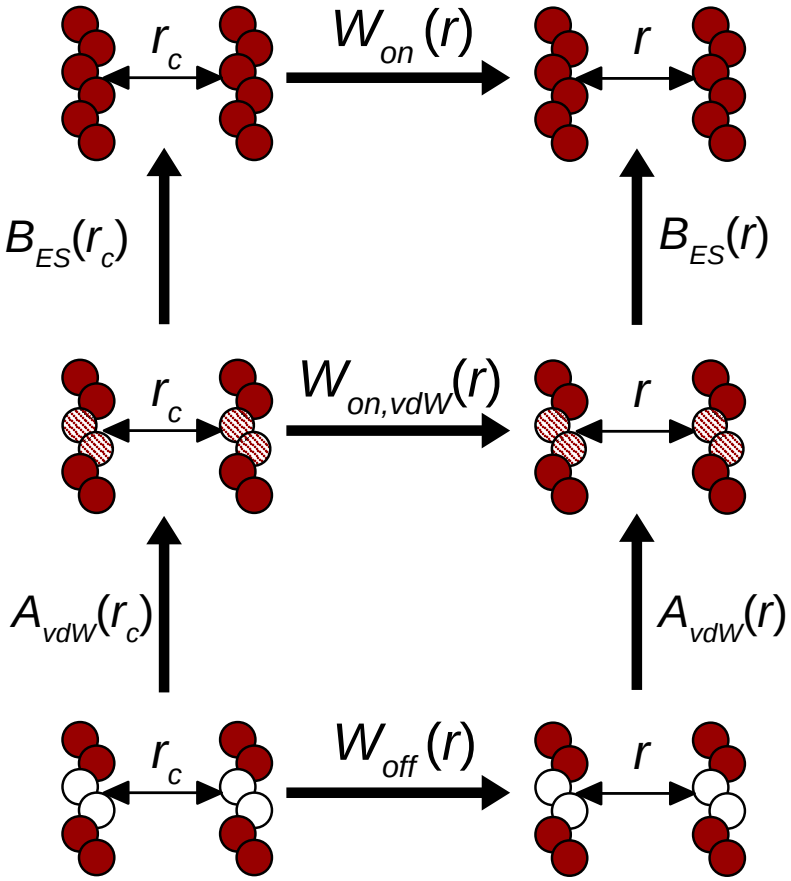


Figure 5.1: Schematic representation of the ECRW method. The CG interaction potentials A_{vdW} and B_{ES} represent the coupling free energy of "switching-on" the respective interaction components between the atoms included in a CG site. Here the two central atoms of a hexamer are merged into one CG site. The coupling free energies are calculated through thermodynamic cycles using the reversible work of moving the two sites of interest from quasi-infinite cutoff distance (r_C) to the finite distance r.

The electrostatic component is approximated using the Coulomb potential with the net charge q_i of the respective CG site i, calculated as the sum over all partial charges q_a of atoms a contained within the site:

$$q_i = \sum_{a \in i} q_a \tag{5.6}$$

$$U_{\mathrm{ES},ij}(r_{ij}) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \tag{5.7}$$

5.2.2 Sampling of CRW Trajectories

In past applications of the CRW method, the sampling of the constraint distance trajectories was performed by molecular dynamics simulations with a constraint algorithm ensuring a constant distance between beads of interest. In this work, we use Monte Carlo sampling to generate the trajectories used for the potential parametrization. The reason for this is that for systems with strong electrostatic interactions, such as the ionic liquids investigated here, we experienced sampling problems in constraint distance MD simulations. Since there are only two molecules in vacuo, highly attractive electrostatic energy between certain atoms can lead to conformations that are "trapped" in a local energy minimum. This hinders a sampling of the whole configuration space and therefore leads to very poor-quality potentials. Using Monte Carlo moves, these local minima can be escaped thereby overcoming sampling issues. Also, the simulations with the Monte Carlo approach tend to converge faster (with respect to the number of steps) than MD simulations performed in the past.

The Monte Carlo algorithm employed here performs two kinds of trial moves:

1. Rotation of the whole molecule around the x-,y- or z-axis (one of which is randomly chosen)
2. Torsional rotation of a molecule around a randomly selected (σ)-bond

After each move, the modified molecule is shifted such that the distance r_{ij} between the relevant CG sites is kept constant. Moves are accepted or rejected using the Metropolis criterion:

$$p_{\mathrm{accept}} = \begin{cases} 1 & \text{if } \Delta U \leq 0 \\ \exp\left(-\frac{\Delta U}{k_\mathrm{B} T}\right) & \text{if } \Delta U > 0 \end{cases}, \tag{5.8}$$

where $\Delta U = U_{\mathrm{new}} - U_{\mathrm{old}}$ is the energy difference between the current and the previous configuration. T and k_B denote the temperature and Boltzmann constant, respectively. After accepting or rejecting the trial configuration, F_C (eq. 5.1) is calculated from the current configuration.

With this Monte Carlo simulation we can efficiently sample spatial orientations and torsional states of the molecules. Bond lengths and angles are kept constant in this approach. We assume their influence on the resulting potential to be negligible.

5.3 Parameterization of the Coarse-Grained Model

5.3.1 Mapping Scheme

In this work three different ionic liquid systems will be simulated: a butyl-methyl-imidazolium cation ($[\mathrm{BMIM}]^+$) is combined with hexafluorophosphate ($[\mathrm{PF_6}]^-$), tetrafluoroborate ($\left[\mathrm{BF_4}\right]^-$) and chloride ($[\mathrm{Cl}]^-$) anions. For $[\mathrm{BMIM}]^+$and $[\mathrm{PF_6}]^-$, we use the all-atom force field published by Bhargava and Balasubramanian [62]. The interaction parameters of $\left[\mathrm{BF_4}\right]^-$are taken from a different force field[55],

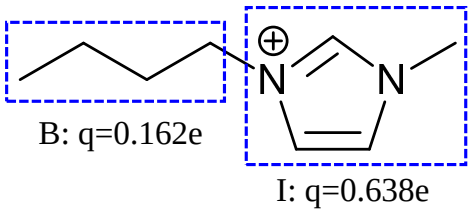


Figure 5.2: Mapping scheme employed for the coarse-graining of the BMIM cation including the net charges of sites B and I as calculated from the atomistic charges.

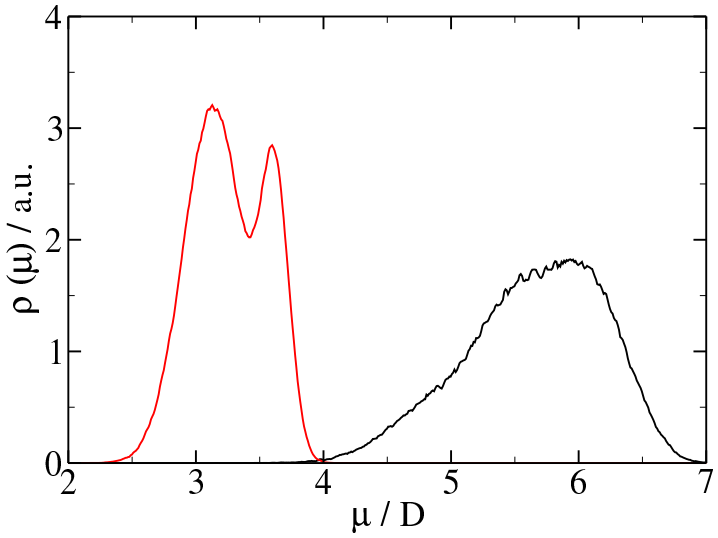


Figure 5.3: Dipole moment distribution of the $[BMIM]^+$cation in atomistic (black) and coarse-grained simulations (red) of $[BMIM]^+[PF_6]^-$. The distribution is almost identical in the systems with $[BF_4]^-$and $[Cl]^-$anions.

while for $[Cl]^-$we use the Lennard-Jones parameters of the OPLS-AA forcefield.[63] The cation and all anions have an absolute charge of 0.8 e.

On the basis of this atomistic representation of $[BMIM]^+[X]^-$, we define a center-of-mass mapping that maps the anions on one site and the cation on two sites. The cation is mapped such that the butyl and methylimidazolium units are contained within one CG site each (see fig. 5.2), named B (butyl) and I (methylimidazolium), respectively. The sum of the atomistic charges yields a total net charge of 0.162 e (B) and 0.638 e (I).

In this work, a two-site center of mass description of the $[BMIM]^+$cation is used. Other mappings that have been proposed for highly polarizable molecules are based on the center of charges.[64] In principle, a CG model based on such a mapping could also be parameterized using the ECRW method.

Most CG models of $[BMIM]^+$-based RTILs use a three site mapping scheme with the intent of better reproducing structural and orientational correlations.[49, 50, 58] The computational advantages of a coarser model (fig. 5.2) of course come at the cost of losing the ability to describe certain properties. While applying total net charges on the mapped beads provides an accurate description of electrostatic bead-bead interactions between ion pairs (shown later), it does not provide a good representation of the cation's average dipole moment (fig. 5.3). Therefore, we expect discrepancies between dielectric properties predicted by the CG and FG models.

5.3.2 Interaction Potentials

Using the Monte Carlo algorithm described above, we perform simulations to generate interaction potentials for $[BMIM]^+[X]^-$. To this end, MC simulations of all CG site pairs in vacuo are performed at constant distances from 0.30 nm to 1.30 nm with a distance increment of 0.02 nm. The cutoff length for the Lennard-Jones potential is set to 1.3 nm, while electrostatic interactions are simulated using the Coulomb potential with infinite cutoff length. Single simulations have a length of 10^6 MC moves and the temperature is set to 500 K. Spatial and torsional rotation each make up for 50% of the trial moves. The maximum attempted rotation angle is 90° for both molecular and torsional rotations. The overall acceptance rates are between 30% and 75% depending on the molecular distance. The force between the two molecules is written every 100 moves.

A relative dielectric permittivity of 1 is used in the MC vacuum simulations. In our earlier publication on ECRW we used the experimental value of ε_r with the intent of mimicking the electrostatic environment of the liquid.[4] We decided not to do so here because this screening is (a) not necessary on long distances since the Coulomb form is reasonably well reproduced by all pairs here, and (b) because it is unrealistic at short distances and a screening here may mask inaccuracies of the point charge approximation by reducing the mismatch between the observed and fitted electrostatic interaction energy.

After integration of the molecular pair force, the potentials A_{vdW} and B_{ES} are obtained. A_{vdW} is fitted using the equation 5.5. The resulting parameters are shown in table 5.1. In fig. 5.6 we compare the data points obtained for the potential of the I-I pair to the resulting fitting curve. The uncertainty of the calculated potential is estimated by calculating A_{vdW} separately using five blocks of configurations representing 200000 MC trial moves and assuming as uncertainty of the mean value at each distance the respective standard deviation between the five potentials curves. The so-obtained uncertainty is approximately equal for all potential curves.

B_{ES} is fitted using the Coulomb potential and the net charge of the respective coarse-grained bead resulting from the mapping scheme employed in the coarse-graining procedure (see Fig. 5.4 and 5.5 for a comparison of the potentials resulting from the ECRW method and the Coulomb potentials used in the CG simulations). The Coulomb function used for fitting function is shifted such that $B_{ES} = 0$ at the cut-off distance. In the CG simulations, this shift is not performed and we simulate the electrostatic interactions using the standard Coulomb potential with the charges in tab. 5.1 for the short range and the P3M long-range method. From this comparison we conclude that the point charge assumption made in the CG model is justified at long distances for almost all interaction pairs.

Deviations on shorter distances occur for almost all interactions. These discrepancies originate from the fact that we represent the (fine-grained) charge distribution by a single charge at the center of mass. At long distances, the monopole contribution dominates and the Coulomb-approximation is well justified. At closer distances, however, multipole contributions to the electrostatic interactions become more important leading to a deviation between the observed effective interaction potential and the Coulomb functional form. This effect is strongest for the I-I pair because the imidazolium group shows the most asymmetric charge distribution and therefore has the largest multipole contribution to the effective electrostatic interaction.

The B-Cl interaction is the interaction that is least well approximated by the Coulomb potential. We expect an attractive interaction but find that the potential shows a repulsive character at short distances. This mismatch can probably be attributed to sampling problems in the ECRW simulations.

In initial bulk simulations for $[BMIM]^+[PF_6]^-$at 500 K, the liquid density was slightly overestimated (see tab. 5.2). An overestimation of mass density has been observed in many CRW models.[3, 4, 43] Herein, we propose a correction to the nonbonded interactions to reduce the mismatch in the density by slightly enhancing the strength of the repulsive interactions between selected pairs of CG sites. We choose this approach, since it has been observed in previous studies[3, 4, 43–45], that CRW models are very well suited to reproduce thermodynamic properties (e.g. the vapor-liquid surface tension), a quality that has been to a large part explained by the attractive part of the CG potentials. A correction acting on

Table 5.1: Non-bonded parameters of the force field used for the CG simulations. The unmodified fitting parameters of A_{vdW} are noted in parentheses.

	ε / kJ mol^{-1}	r_0 / nm	k_1 / nm^{-1}	k_2 / nm^{-1}	q / e
B-B	1.87	0.671	6.97	10.8	0.162
B-I	2.67	0.601	5.87	10.1	
I-I	3.25	0.567	6.00 (5.39)	9.12	0.638
B-PF_6	2.90	0.573	10.3	7.98	
I-PF_6	3.44	0.570 (0.556)	8.50 (6.91)	10.6	
PF_6-PF_6	3.16	0.571	12.3	11.4	-0.8
B-BF_4	1.98	0.531	12.0	7.56	
I-BF_4	2.68	0.550 (0.536)	10.0 (6.92)	10.7	
BF_4-BF_4	1.56	0.526	12.0 (10.2)	12.3	-0.8
B-Cl	2.18	0.447	15.0 (13.0)	7.96	
I-Cl	2.06	0.500 (0.486)	9.70 (8.61)	12.1	
Cl-Cl	1.23	0.378	18.5	12.7	-0.8

the attractive part of the potential might impact this particular quality of the CG potential in a disadvantageous manner and we therefore decide to tune the repulsive part. After fitting the vdW component of the ECRW potentials with the modified Morse potential we can use the repulsive stiffness parameter k_1 to systematically increase the repulsiveness of the potentials and thereby reduce the equilibrium density of the liquid. For this modification, we select pairs with a comparatively large mismatch in the pair correlation function and modify the corresponding k_1 (see tab. 5.1). Further, r_0 of the I-PF_6, I-BF_4, and I-Cl interaction potentials is increased by 0.014 nm. We stop the modification of the potential parameters if the model reproduces the density with less than 5% relative error and approximately reproduces the position of the first peak of the RDF.

An overview of A_{vdW} and the potential used in the model is given in fig. 5.7 for all pairs, where such modifications have been performed. Using these modified potentials will lead to a more accurate reproduction of the density, without drastically changing the shape of the interaction potentials.

The intramolecular degree of freedom that is remaining within the $[\mathrm{BMIM}]^+$ion in the CG representation (i.e. the bond length between the B and I site) is parameterized using the Boltzmann inversion method. An MD trajectory of 100 ns is generated using one FG molecule in vacuo at 500 K, using a Langevin thermostat with a time constant of 1 ps. All nonbonded interactions between atoms are excluded from this simulation. The trajectory is mapped using the mapping scheme introduced above and the distance distribution $\rho(r)$ between the B and I sites is extracted from the mapped configurations. From this distribution the bonded interaction potential is calculated through:

$$U_{\mathrm{bond}}(r) = -k_{\mathrm{B}}T \ln\left(\frac{\rho(r)}{r^2}\right) + C, \tag{5.9}$$

where C is chosen such that $\min(U_{\mathrm{bond}}) = 0$. The resulting potential is shown in fig. 5.8.

5.4 Computational Detail

For all systems, we conduct bulk MD simulations in an NPT-ensemble 1 bar and 5 temperatures between 300 K and 500 K. We generate 2 ns of trajectory starting from equilibrated initial structures using a time

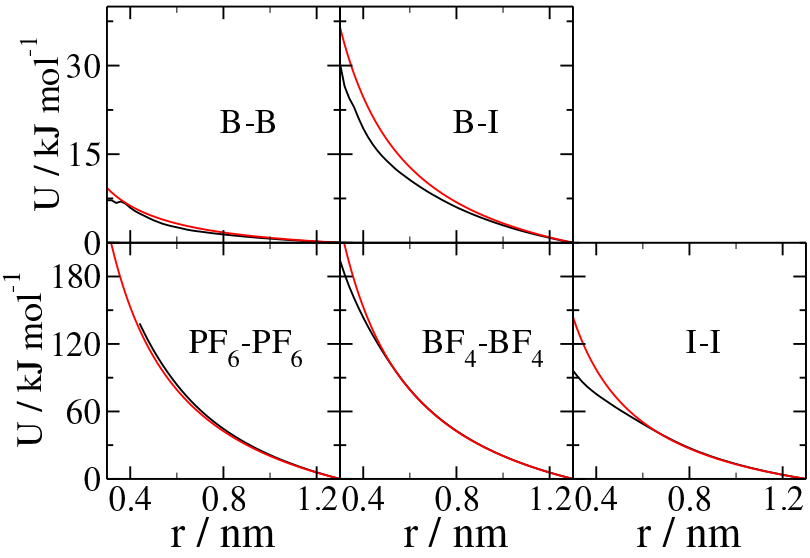


Figure 5.4: Comparison of the electrostatic component of the CRW potential B_{ES} (black) and the Coulomb potential (shifted to zero at the cut-off distance of 1.3 nm) using the net charges of the CG sites (red) for all pairs with repulsive electrostatic interaction.

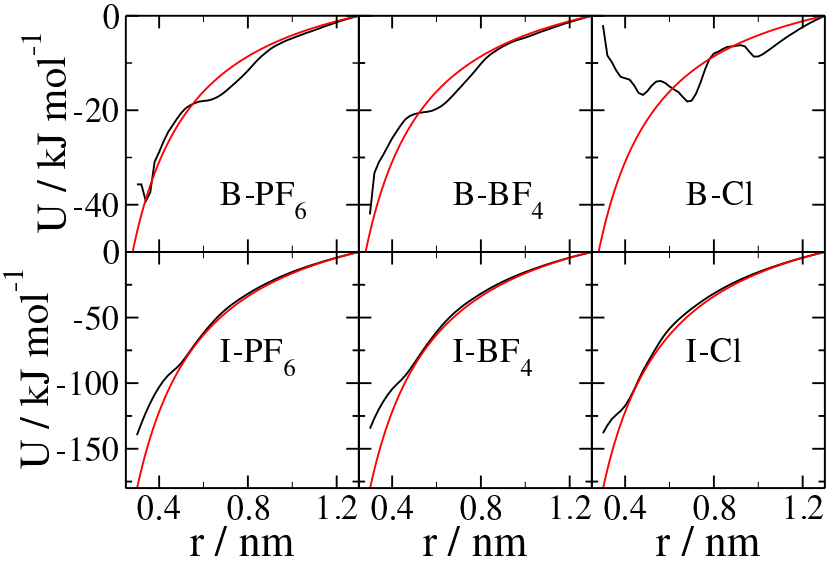


Figure 5.5: Comparison of the electrostatic component of the CRW potential B_{ES} (black) and the Coulomb potential (shifted to zero at the cut-off distance of 1.3 nm) using the net charges of the CG sites (red) for all pairs with attractive electrostatic interaction.

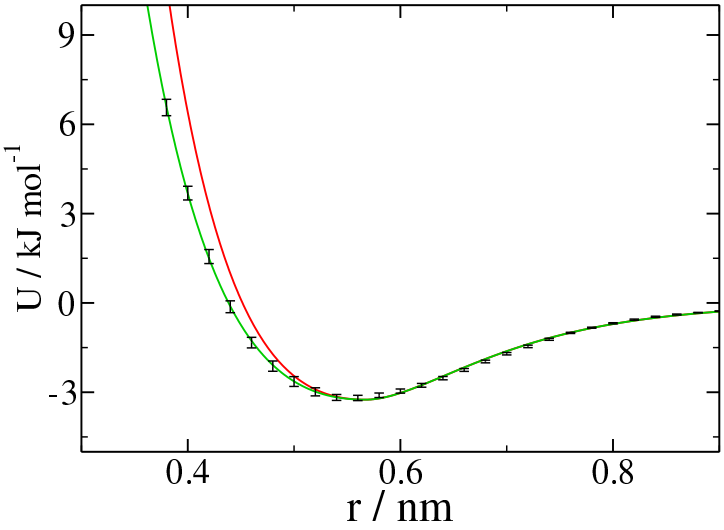


Figure 5.6: Comparison of the calculated CRW potential A_{vdW} (black with error bars) and the interaction potential obtained from the initial fit (green) and that used in the CG simulation (red) for the I-I site pair.

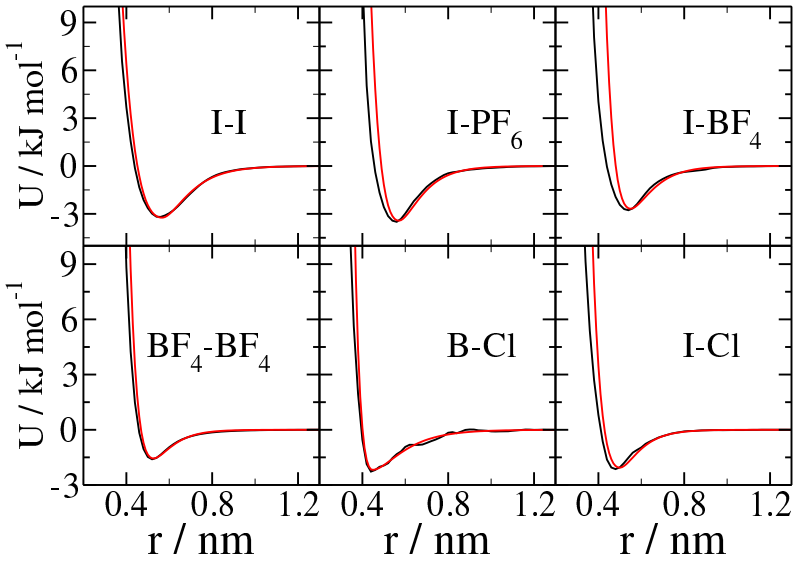


Figure 5.7: Comparison of the calculated CRW potential A_{vdW} (black) and the interaction potential used in the CG simulation (red) for all pairs, where modifications to the original fit have been performed.

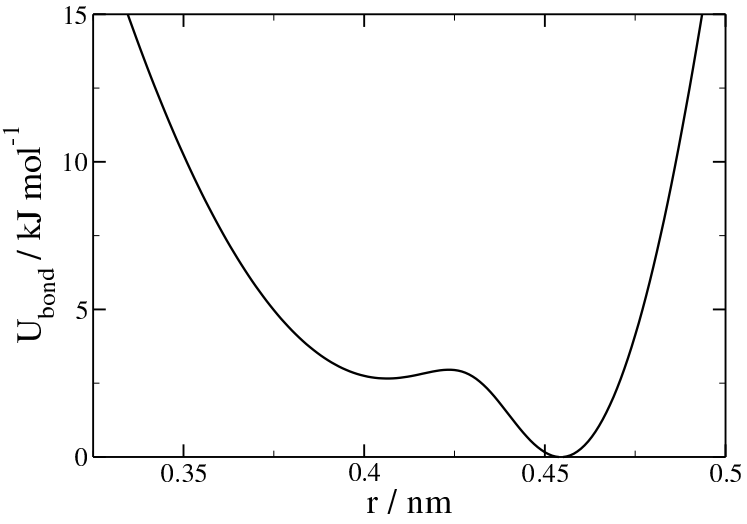


Figure 5.8: Bonded potential used to model the B-I bond.

Table 5.2: Resulting liquid-vapor surface tensions and mass densities of the optimized and unmodified CG models at 500 K

	ρ / kg m^{-3}			γ_{lv} / mN m^{-1}	
system	AA	CG	$CG_{unmodified}$	AA	CG
$[BMIM]^+[PF_6]^-$	1216	1221	1386	28.5±7.5	28.0±0.3
$[BMIM]^+[BF_4]^-$	1035	1041	1287	25.7±1.8	26.5±0.4
$[BMIM]^+[Cl]^-$	1011	1071	1154	28.4±2.6	28.3±0.4

step of 1 fs. Cut-off distances are chosen at 1.3 nm. The Nosé-Hoover thermostat[65] and Parrinello-Rahman barostat[66] are used with respective time constants of 1 and 5 ps. The atomistic simulations were conducted using Gromacs version 4.6.7[67], while the CG simulations were performed using LAMMPS.[68] The long-range part of the electrostatics is simulated using the Particle-Mesh-Ewald[69] (PME, AA systems) and Particle-Particle-Particle-Mesh[70] (P3M, CG systems) algorithms. In the atomistic simulations long-range corrections for the Lennard-Jones potentials are applied to pressure and energy. In the CG simulations, the bond distance potential energy is implemented using a lookup table with 1001 entries between 0 nm and 1 nm obtained from spline interpolation between the data points obtained from the parametrization (see fig. 5.8).

In addition to the NPT bulk simulations, NVT simulations are performed to study the behavior of the liquid-vapor interface. The initial configuration of these simulations is obtained by extending the simulation box of an equilibrated configuration by a factor of 2 in the z-direction. To this end, the final configuration from the NPT bulk liquid simulation is taken and the box is extended by a factor of 2 in the z-direction. Simulations of the liquid-vapor interface are performed using NVT molecular dynamics with identical settings as for the NPT bulk simulations described above and a simulation length of 50 ns.

5.5 Simulation Results and Discussion

Table 5.3: Isothermal compressibility κ, relative dielectric permittivity ε_r and finite system Kirkwood factor G_k calculated from simulations with atomistic and CG models at 500 K

	κ / 10^{-9} Pa^{-1}		ε_r		G_k	
system	AA	CG	AA	CG	AA	CG
$[BMIM]^+[PF_6]^-$	0.85	1.18	6.7	4.0	1.3	2.1
$[BMIM]^+[BF_4]^-$	0.98	1.38	10.5	4.4	1.8	2.3
$[BMIM]^+[Cl]^-$	0.64	1.07	10.1	4.6	1.6	2.0

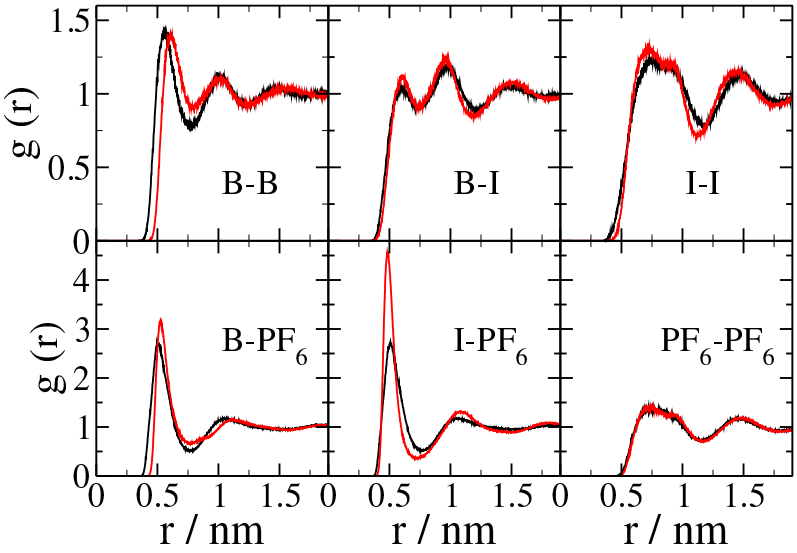


Figure 5.9: Radial distribution functions of all site pairs of $[BMIM]^+[PF_6]^-$ as calculated from the atomistic (black) and coarse-grained (red) simulations.

5.5.1 Properties of the Liquid Phase

The mass densities resulting from the NPT simulations are listed in table 5.2. We observe that the densities from the CG simulation match the atomistic reference for $[BMIM]^+[PF_6]^-$ and $[BMIM]^+[BF_4]^-$. This was intended by the modification of the potential parameters of the CG site pairs. On the other hand, the density of $[BMIM]^+[Cl]^-$ is roughly 6 % higher in the CG model as compared to the atomistic model. Reducing the density mismatch further with the modification of the cation-anion vdW interactions is only possible at the cost of an even stronger mismatch in the RDF, which we intend to avoid.

Figures 5.9-5.11 show the radial distribution functions $g(r)$ for all systems in the atomistic and CG representation. For $[BMIM]^+[PF_6]^-$ the structure is well reproduced (see fig. 5.9). The CG model is capable to reproduce the $g(r)$ for the pairs I-I, B-I, and PF_6-PF_6 almost quantitatively. For the remaining pairs, we can observe small deviations with respect to the excluded volumes (larger for B-B and B-PF_6) which is probably due to the mapping of an almost linear chain on a spherical CG site and the resulting underestimation of orientation effects.[32] Further, the first peak in the $g(r)$ of I-PF_6 is much higher in the CG simulation than in the atomistic reference. We attribute this to the fact that the charge of the I bead is localized in one point in the CG model as opposed to the charge distribution over multiple atoms in the atomistic model. An analogous argument can be made for the B-PF_6 RDF, which suffers from a similar mismatch.

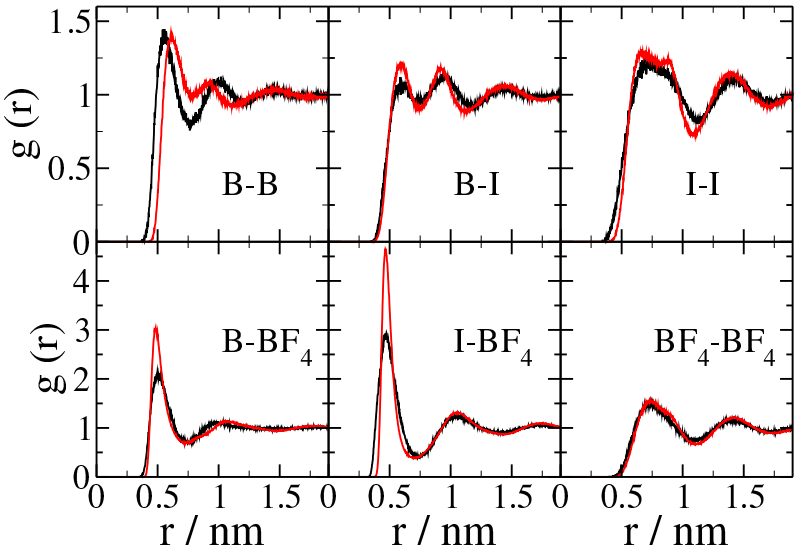


Figure 5.10: Radial distribution functions of all site pairs of $[\mathrm{BMIM}]^{+}[\mathrm{BF_4}]^{-}$ as calculated from the atomistic (black) and coarse-grained (red) simulations.

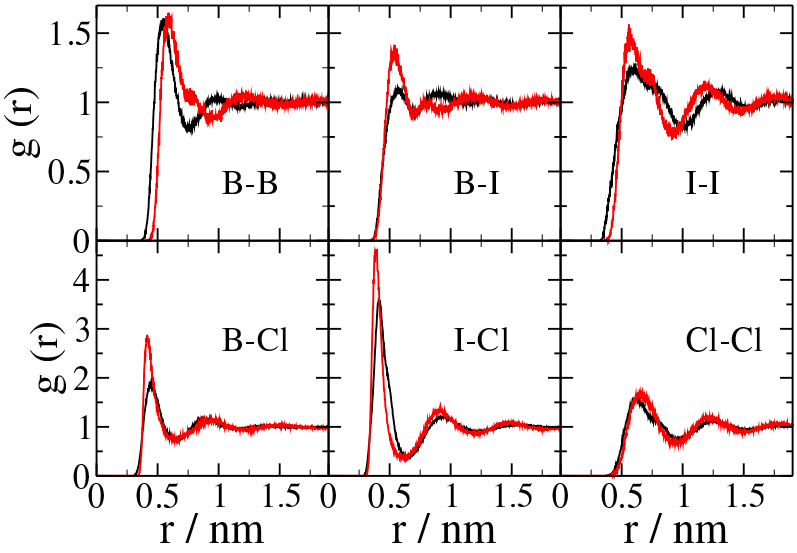


Figure 5.11: Radial distribution functions of all site pairs of $[\mathrm{BMIM}]^{+}[\mathrm{Cl}]^{-}$ as calculated from the atomistic (black) and coarse-grained (red) simulations.

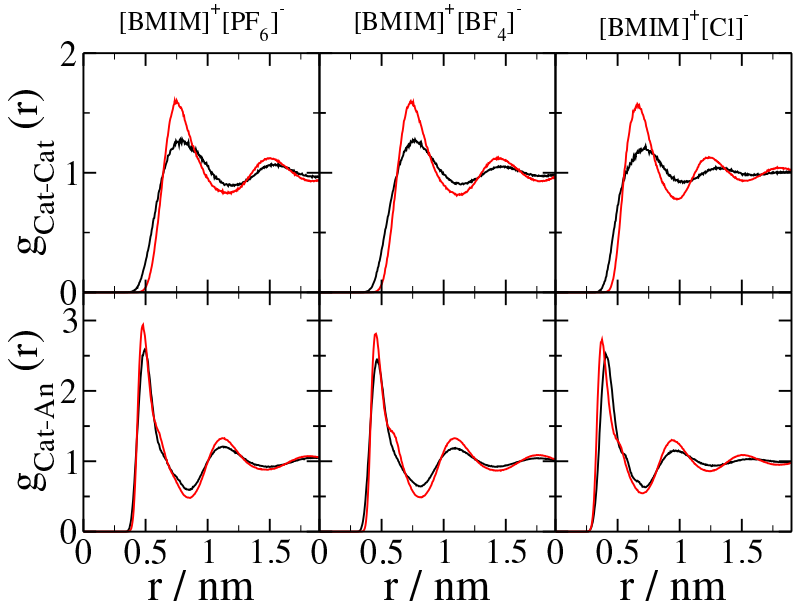


Figure 5.12: Radial distribution functions of the molecular centers of mass. Distributions of cation-cation pairs and cation-anion pairs for all systems are shown in the upper and lower half, respectively, for the atomistic (black) and CG systems (red).

Table 5.4: Coordination numbers of the first solvation shell of all CG site pairs.

	$[BMIM]^+[PF_6]^-$		$[BMIM]^+[BF_4]^-$		$[BMIM]^+[Cl]^-$	
pair	AA	CG	AA	CG	AA	CG
B-B	4.59	4.56	4.98	5.04	5.21	5.31
B-I	3.33	3.35	3.73	3.85	3.60	4.00
I-I	17.5	17.1	16.5	16.3	14.2	11.0
B-$[X]^-$	5.09	5.22	5.01	5.18	4.82	5.38
I-$[X]^-$	5.10	5.04	5.23	5.07	5.70	5.32
$[X]^-$-$[X]^-$	15.6	15.9	16.2	16.3	14.6	15.9

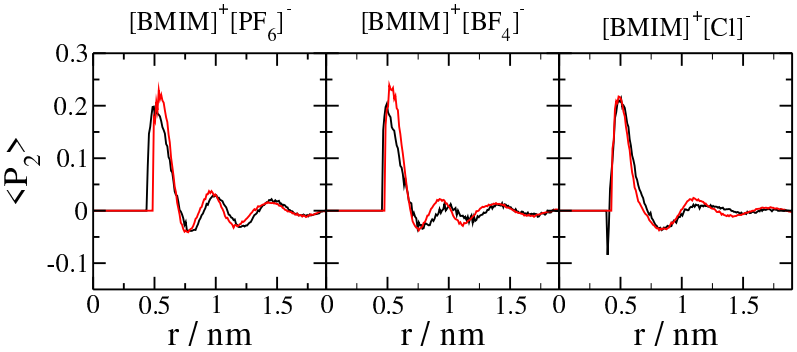


Figure 5.13: Orientational correlations between the $[BMIM]^+$cations for the atomistic (black) and CG models (red).

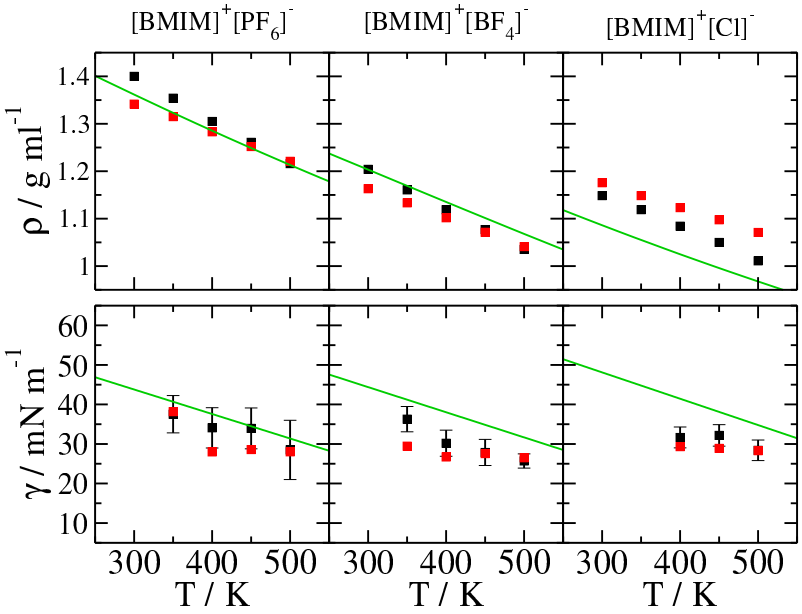


Figure 5.14: Temperature transferability of mass density (upper half) and liquid-vapor surface tension (lower half). Black symbols denote results from atomistic simulations, red symbols denote results from CG simulations and green lines depict an extrapolation based on experimental values of ρ[71, 72] and γ_{lv}[73].

For $[\mathrm{BMIM}]^+[\mathrm{BF_4}]^-$ similar conclusions can be drawn (see fig. 5.10). The reproduction of the $g(r)$ is good by standards of other CRW models. However, it is less accurate than the reproduction for the system containing $[\mathrm{PF_6}]^-$. Notable effects in the $[\mathrm{PF_6}]^-$system are also observed here: namely an increased excluded volume for B-B and an increased first peak for I-$\mathrm{BF_4}$.

The CG model for $[\mathrm{BMIM}]^+[\mathrm{Cl}]^-$(see fig. 5.11) does not perform to the same degree as the two other models. The $g(r)$ is only reproduced to quantitative extent for the Cl-Cl pair, the potential of which is identical in the CG and atomistic models. For other pairs a qualitative reproduction of characteristic features is achieved. Most notably, the $g(r)$ of the B-I pair considerably deviates from the atomistic reference. This mismatch is probably caused by the (simplified) modeling of the electrostatic interactions by point charges in the CG model, which is not able of reproducing all multipole contributions to the ES interactions in detail. Figure 5.5 shows that the deviation from point charge behavior is largest for the B-Cl site pair. The overestimation of attractive electrostatic forces is countered, to a certain degree, by the modification of the effective Van-der-Waals interactions, but still we anticipate that the balance between the site-site interactions is somewhat ill-represented in the resulting CG model. The fact that the $[\mathrm{Cl}]^-$ion shows the highest charge density among the studied anions adds to this effect, since for this system the electrostatic forces dominate the pair interactions. This contribution, however, is probably the interaction component which is modeled less accurately than the vdW component by the ECRW method leading to a better performance of the ECRW models for the systems in which the anions have a lower charge density.

The coordination number of the first coordination shell is shown in table 5.4 for all site pairs. These results support the finding that the structure is well reproduced by the CG models of $[\mathrm{BMIM}]^+[\mathrm{PF_6}]^-$and $[\mathrm{BMIM}]^+[\mathrm{BF_4}]^-$, and to a lesser degree also by $[\mathrm{BMIM}]^+[\mathrm{Cl}]^-$.

A comparison of the RDFs of the molecular centers of mass for cations and anions (fig. 5.12) gives a further interesting insight into the differences in the liquid structure between the two models. In these RDFs, a mismatch in the cation-cation distribution is observed: the first peak of these distributions is higher and narrower for all CG systems studied. This is especially interesting since none of the site-site RDFs (Figures 5.9-5.11) shows a comparably large mismatch. On the other hand, the cation-anion RDFs

are far better reproduced, despite larger mismatches in the RDFs of the single CG site pairs. For the anion-anion RDFs (see figures 5.9-5.11) the distributions calculated from CG simulations reproduce the atomistic reference.

An explanation for this overstructuring of the cation-cation pair correlations may be found in the rather coarse two-site mapping scheme chosen for the cation. Another cause for this mismatch may be the simplification of the electrostatic interactions. While the assumption of a point charge at the center of mass is a fair assumption for the highly rigid and spherically symmetric anions, it is a very coarse approximation of the cation, which is more flexible and has a smeared out distribution of the charge within the molecule. The cation in the CG representation will necessarily be more rigid in terms of the charge distribution, and this fact reflects in more pronounced minima and maxima in the cation-cation RDF.

In figure 5.13 we show the orientation of the $[\mathrm{BMIM}]^+$cation as a function of the COM distance:

$$\langle P_2(r)\rangle = \frac{1}{2}\langle 3\cos^2(\theta) - 1\rangle, \tag{5.10}$$

where θ is the angle between the B-I distance vectors of two molecules at distance r. We find that the orientational structure is reproduced not quantitatively but in reasonable agreement for all of the investigated systems.

The temperature dependence of the mass density is compared in figure 5.14. All CG models slightly underestimate the thermal expansion coefficient in comparison with the atomistic reference. Still, the mass densities are reproduced in reasonable agreement (deviation of less than 5%) even at the lowest temperature which is 200 K lower than the temperature at which the potentials were parameterized. We note further that the agreement between simulation data and experimental results is best for $[\mathrm{BMIM}]^+[\mathrm{PF_6}]^-$since the FG reference model has been parameterized for this particular system. For $[\mathrm{BMIM}]^+\left[\mathrm{BF_4}\right]^-$and $[\mathrm{BMIM}]^+[\mathrm{Cl}]^-$the CG model performs worse with respect to the experimental data but this is mainly due to shortcomings of the FG models.

Additional bulk properties are shown in table 5.3. Here we present the isothermal compressibility κ:

$$\kappa = \frac{\langle V^2\rangle - \langle V\rangle^2}{k_\mathrm{B}T\langle V\rangle}, \tag{5.11}$$

where V denotes the total volume, T the temperature and, k_B the Boltzmann constant. The CG models overestimate κ which is in agreement with CRW models reported for other systems,[4, 44]and is casued by slightly softer interactions in the CG model.

We also calculate the relative dielectric permittivity ε_r:[74]

$$\varepsilon_\mathrm{r} = 1 + \frac{\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2}{3\varepsilon_0 V k_B T}, \tag{5.12}$$

using the permittivity of the vacuum ε_0 and the total electric dipole moment of the simulation box $\mathbf{M}$. The CG models all underestimate the atomistic values for ε_r. As discussed in section 5.3, the mapping scheme of the cation underestimates the molecular dipole moment μ with the result that also ε_r is underestimated.

We further calculate the finite size Kirkwood factor G_k which measures the strength of orientational correlation of one dipole with all other dipoles in the system:[75]

$$G_k = \frac{\langle \mathbf{M}^2\rangle - \langle \mathbf{M}\rangle^2}{N_\mathrm{mol}\langle \mu^2\rangle}, \tag{5.13}$$

where the total dipole moment fluctuation is normalized by the number of molecules N_mol and the mean square dipole moment of the molecules $\langle \mu^2\rangle$. G_k is overestimated by all CG models indicating stronger correlations between dipole moments than in the FG systems. This finding is consistent with the overstructuring observed in the molecular RDF (fig. 5.12) and, to a lesser degree, the orientational correlations between pairs (fig. 5.13).

Table 5.5: Mass density at $z = 0$ calculated from the density profiles (fig. 5.15)

	ρ / kg m^{-3}	
$[BMIM]^+[PF_6]^-$	1201	1218
$[BMIM]^+[BF_4]^-$	1020	1044
$[BMIM]^+[Cl]^-$	1002	1070

5.5.2 Properties of the Liquid-Vapor Interface

From the simulations of the liquid-vapor interface we calculate the liquid-vapor surface tension γ_{lv} using the well-established relation:

$$\gamma_{lv} = \frac{L_z}{2}\left\langle p_{zz} - \frac{p_{xx} + p_{yy}}{2}\right\rangle, \tag{5.14}$$

where L_z is the length of the box parallel to the interface normal vector (in this case the z-axis) and p_{ii} denotes the diagonal elements of the stress tensor. Table 5.2 shows that the values for γ_{lv} calculated from CG simulations match the atomistic reference values for all RTILs within error bars. This result shows that ECRW-based potentials for the ionic liquid systems are transferable to liquid-vapor interfaces, a result that has been observed for other CRW CG models.[3, 4, 44]

Finite-size effects can play a critical role in the calculation of the liquid-vapor surface tension of ionic liquids.[76] To check our results for consistency, we perform an additional simulate with the atomistic $[BMIM]^+[Cl]^-$system, with a larger box (2048 ions instead of 512). The calculated γ_{lv} from this simulation is 28.1 mN m^{-1}, which is identical within error bars to the simulation result with the smaller box.

The temperature dependence of γ_{lv} is shown in figure 5.14. This dependence, like other properties, is best reproduced in comparison with experimental data for the $[BMIM]^+[PF_6]^-$and $[BMIM]^+[BF_4]^-$systems, while the variation of γ_{lv} with T is underestimated for $[BMIM]^+[Cl]^-$.

In addition to γ_{lv}, we use the results from the liquid-vapor interface simulation to analyze the density profile along the interface for all systems. The densities of the respective anion and cation molecules are shown in fig. 5.15. Much like in the case of the liquid structure, we can observe that the CG models of $[BMIM]^+[PF_6]^-$and $[BMIM]^+[BF_4]^-$reproduce the atomistic density distribution of both the cation and anion well with only small deviations. For the $[BMIM]^+[Cl]^-$the deviations of the CG simulations with respect to the atomistic system are larger, as might be expected from the deviations observed for the bulk density. Still the CG model is capable of a qualitative reproduction of the density profile for both ion types. The bulk mass densities $\rho(z = 0)$ of the three system are calculated from the density profiles. We find that the values are consistent with the results obtained from the NpT simulation of the bulk liquid (tab 5.5). Additionally, we calculate the potential of mean force of the cations and anions as a function of z (fig. 5.16):

$$PMF(z) = -k_B T \log \frac{\rho(z)}{\rho(0)} \tag{5.15}$$

5.5.3 Dynamic Properties and Speed-Up Rates

The self-diffusion coefficient D can be calculated from the slope of the mean squared displacement $\langle \Delta r^2 \rangle$ using the Einstein relation:

$$D = \lim_{t\to\infty} \frac{1}{6t}\left\langle \Delta r^2(t)\right\rangle \tag{5.16}$$

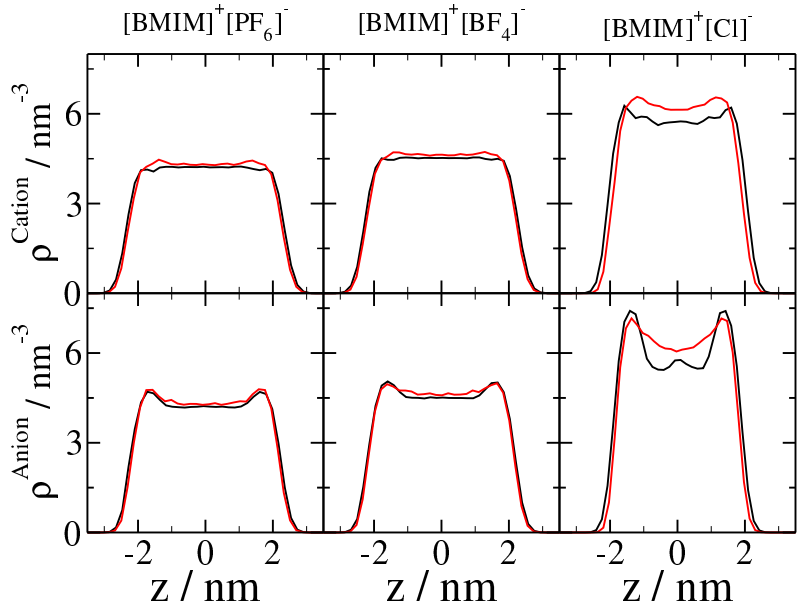


Figure 5.15: Molecular number density along the z-axis of a half empty box for all ionic liquids. We compare the atomistic reference (black) to the CG simulation results (red) for $[\mathrm{BMIM}]^+$(upper half) and the respective anion (lower half).

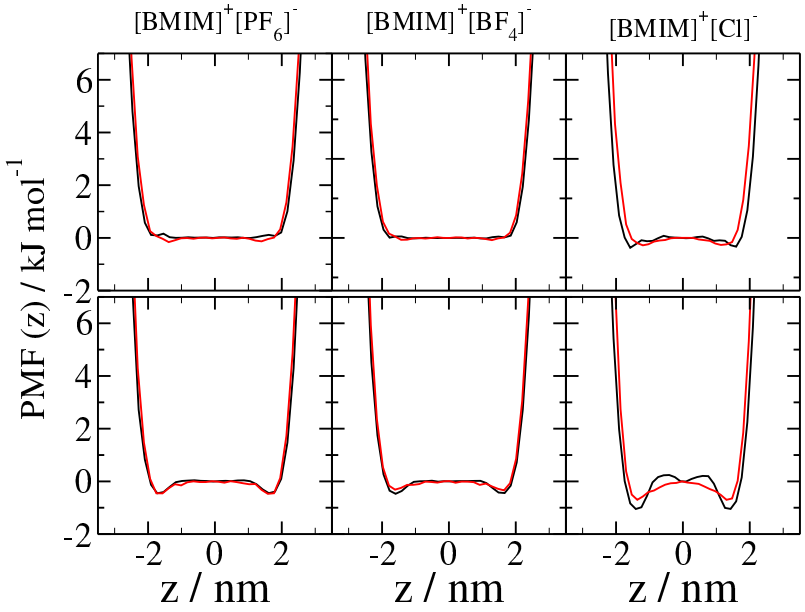


Figure 5.16: Potential of mean force of the ions along the z-axis of a half empty box for all ionic liquids. We compare the atomistic reference (black) to the CG simulation results (red) for $[\mathrm{BMIM}]^+$(upper half) and the respective anion (lower half).

Table 5.6: Self-diffusion coefficients and ionic conductivities of all investigated systems. Experimental values are based on the model in ref. [78]

System		D / 10^{-5} cm^2 s^{-1} AA	CG	Exp.	Speed-up D_{CG}/D_{AA}	λ / S m^{-1} AA	CG	Speed-up $\lambda_{CG}/\lambda_{AA}$
$[BMIM]^+[PF_6]^-$	Cation	1.05	1.17	1.46	1.11	27.5	35.7	1.3
	Anion	0.65	1.24	1.86	1.91			
$[BMIM]^+[BF_4]^-$	Cation	1.24	1.55	1.44	1.25	96.6	102	1.1
	Anion	1.23	1.74	2.89	1.41			
$[BMIM]^+[Cl]^-$	Cation	0.55	1.74		3.16	50.2	104	2.1
	Anion	0.74	2.31		3.12			

In addition we calculate the ionic conductivity λ of the ionic liquids using the Green-Kubo relation:[77]

$$\lambda = \frac{1}{3Vk_BT}\int_0^\infty \langle \mathbf{J}(t)\cdot\mathbf{J}(0)\rangle\, dt, \tag{5.17}$$

where $\mathbf{J}(t) = e\sum_{i=0}^{n_{ion}} z_i \mathbf{v}_i(t)$ is the current calculated from the ions' respective velocities and charges.

The self-diffusion coefficients and ionic conductivities calculated from the trajectories of the liquid system are listed in table 5.6.

In general, the diffusion coefficients are higher in the CG system. An acceleration of dynamics is common in systematically coarse-grained models that use the standard Newtonian equation of motion. It is linked to the softer potential energy surface of the CG model in comparison to the fine-grained representation. The speed-up rate is different for all investigated systems and also differs between anions and cations. For a description of the dynamical behavior that is consistent with the fine-grained reference model, CG models have to be parameterized using a coarse-graining algorithm that accounts for the dissipative interactions in the CG representation.[79–83] However, such an extensive parametrization would go beyond the scope of this work and a dynamically consistent model is not required in applications that focus on modeling the thermodynamics correctly.

The speed-up rates of the diffusion coefficients are small for the models with $[PF_6]^-$and $\left[BF_4\right]^-$anions and larger for the system with the $[Cl]^-$anion. This finding correlates with the quality of reproduction of the liquid structure.

A similar trend is observed for the ionic conductivity: the speed-up is small for the systems with $[PF_6]^-$and $\left[BF_4\right]^-$, whereas for the $[Cl]^-$system a larger speed-up is observed. A comparison to experimental data is not undertaken for this particular quantity, because the atomistic models used for the reference simulations do not reliably reproduce the experimental ionic conductivities of the investigated compounds.

5.6 Conclusions

In this article, we applied the extended conditional reversible work introduced in an earlier publication[4] to systems of ionic liquids. To this end, we have parameterized CG models for the $[BMIM]^+$cation in combination with $[PF_6]^-$,$\left[BF_4\right]^-$, and $[Cl]^-$anions. The cation is mapped onto two beads while for the anions a single bead representation is used. Small modifications are performed to slightly enhance the repulsive character of the interaction potentials. These modifications do not significantly change the general shape or the attractive part of the potential. The so-obtained ECRW models, which describe the interaction between CG sites by means of effective Van-der-Waals and electrostatic components, are capable of reproducing the liquid bulk density and liquid structure of $[BMIM]^+[PF_6]^-$and $[BMIM]^+\left[BF_4\right]^-$well; for $[BMIM]^+[Cl]^-$, however, the model is less predictive.

The reason for the larger disagreement of the $[BMIM]^+[Cl]^-$CG model with respect to the liquid structure is the high charge density of the $[Cl]^-$anion. In contrast to the $[PF_6]^-$and $\left[BF_4\right]^-$ions, the charge of the $[Cl]^-$ion is not spread over multiple atoms. Therefore, the pair interactions between cation sites and the anion are dominated by electrostatics and deviations of the actual electrostatic interactions from the assumed Coulomb behavior lead to an overall larger modeling error.

The liquid-vapor surface tension has been calculated for the three IL systems. Here we find that for all three systems γ_{lv} is quantitatively reproduced within error bars at 500 K. This finding is in accordance with the results obtained for other systems modeled with CRW-based potentials[3, 4, 44] and illustrates the state-point transferability of these models. A reliable reproduction of the liquid-vapor interface is important especially in the context of modeling wetting phenomena. Further, we find that the ECRW models reproduce the temperature transferability of the mass density and liquid-vapor surface tension reasonably well up to temperatures that are 150 K lower than the temperature of the force field parametrization.

We demonstrate here a method by which CG models of molecules with ionic character can be parameterized in a straightforward and computationally cheap manner from an existing atomistic force field.

The two-site center of mass mapping chosen for the $[BMIM]^+$cation does not give an accurate description of the FG dipole moment. Therefore, dielectric properties of the bulk ionic liquids are not reproduced by the CG models parameterized in this work. The choice of a different mapping, which could be based on centers of charge or contain additional CG sites, might lead to a more accurate description of the dielectric properties.

The self-diffusion coefficients have been studied and we can observe an acceleration of the diffusional dynamics in all studied systems in comparison with the all-atom models. Here again the $[BMIM]^+[Cl]^-$system stands out. The speed-up for this system is larger, while in the other two systems the diffusion coefficients are increased by a lesser degree.

The finding, that models with a strong ionic component can be parameterized using a method based on sampling molecule pairs in vacuo is interesting for further developments of CG models. Since the parametrization of CG models with the ECRW method is computationally inexpensive, it might e.g. be used to explore different mapping schemes and the manner in which the choice of mapping influences the thermodynmmics of the CG model. In addition, it has been shown that the ECRW models reproduce thermodynamics of a fine-grained reference properties well, especially w.r.t. the liquid-vapor surface tension. This makes the resulting potentials suitable e.g. for the simulation of wetting phenomena[84, 85]. Another possibility of further development lies in the inclusion of DPD pair friction to model also dynamic properties of the coarse-grained representation in agreement with the fine-grained reference.[80, 83]

5.7 References

1. Peter, C. & Kremer, K. Multiscale simulation of soft matter systems. *Faraday Discussions* **144.** 00053, 9 (2010).
2. Brini, E. *et al.* Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter* **9.** 00021, 2108 (2013).
3. Brini, E., Marcon, V. & Van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics* **13.** 00022, 10468 (2011).
4. Deichmann, G. & Van der Vegt, N. F. A. Conditional Reversible Work Coarse-Grained Models of Molecular Liquids with Coulomb Electrostatics – A Proof of Concept Study on Weakly Polar Organic Molecules. *Journal of Chemical Theory and Computation* **13,** 6158–6166 (2017).
5. Wang, Y., Noid, W. G., Liu, P. & Voth, G. A. Effective force coarse-graining. *Physical Chemistry Chemical Physics* **11,** 2002 (2009).

6. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128,** 244114 (2008).

7. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24.** 00428, 1624–1636 (2003).

8. Lyubartsev, A. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52.** 00390, 3730–3737 (1995).

9. Ganguly, P., Mukherji, D., Junghans, C. & Van der Vegt, N. F. A. Kirkwood–Buff Coarse-Grained Force Fields for Aqueous Solutions. *Journal of Chemical Theory and Computation* **8.** 00013, 1802–1807 (2012).

10. Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129,** 144108 (2008).

11. Mullinax, J. W. & Noid, W. G. A Generalized-Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials †. *The Journal of Physical Chemistry C* **114,** 5661–5674 (2010).

12. McCarty, J., Clark, A. J., Copperman, J. & Guenza, M. G. An analytical coarse-graining method which preserves the free energy, structural correlations, and thermodynamic state of polymer melts from the atomistic to the mesoscale. *The Journal of Chemical Physics* **140,** 204913 (2014).

13. Johnson, M. E., Head-Gordon, T. & Louis, A. A. Representability problems for coarse-grained water potentials. *The Journal of Chemical Physics* **126,** 144509 (2007).

14. Wagner, J. W., Dama, J. F., Durumeric, A. E. P. & Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *The Journal of Chemical Physics* **145,** 044108 (2016).

15. Dunn, N. J. H., Foley, T. T. & Noid, W. G. Van der Waals Perspective on Coarse-Graining: Progress toward Solving Representability and Transferability Problems. *Accounts of Chemical Research* **49,** 2832–2840 (2016).

16. Hess, B., Holm, C. & Van der Vegt, N. Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Chem. Phys.* **124,** 164509 (2006).

17. Carbone, P., Varzaneh, H. A. K., Chen, X. & Müller-Plathe, F. Transferability of coarse-grained force fields: The polymer case. *The Journal of Chemical Physics* **128,** 064904 (2008).

18. Qian, H.-J. *et al.* Temperature-Transferable Coarse-Grained Potentials for Ethylbenzene, Polystyrene, and Their Mixtures. *Macromolecules* **41,** 9919–9929 (2008).

19. Mognetti, B. M. *et al.* Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide. *J. Chem. Phys.* **128,** 104501 (2008).

20. Allen, E. C. & Rutledge, G. C. Coarse-grained, density dependent implicit solvent model reliably reproduces behavior of a model surfactant system. *The Journal of Chemical Physics* **130,** 204903 (2009).

21. Mognetti, B. M. *et al.* Coarse-grained models for fluids and their mixtures: Comparison of Monte Carlo studies of their phase behavior with perturbation theory and experiment. *The Journal of Chemical Physics* **130,** 044101 (2009).

22. Mullinax, J. W. & Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *The Journal of Chemical Physics* **131,** 104110 (2009).

23. Villa, A., Peter, C. & Van der Vegt, N. F. A. Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water. *J. Chem. Theory Comput.* **6,** 2434–2444 (2010).

24. Cao, F. & Sun, H. Transferability and Nonbond Functional Form of Coarse Grained Force Field – Tested on Linear Alkanes. *Journal of Chemical Theory and Computation* **11,** 4760–4769 (2015).

25. Krishna, V., Noid, W. G. & Voth, G. A. The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. *The Journal of Chemical Physics* **131,** 024103 (2009).

26. Lu, L. & Voth, G. A. The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials. *The Journal of Chemical Physics* **134,** 224107 (2011).

27. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *The Journal of Chemical Physics* **140,** 224104 (2014).

28. Dunn, N. J. H. & Noid, W. G. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *The Journal of Chemical Physics* **143,** 243148 (2015).

29. Foley, T. T., Shell, M. S. & Noid, W. G. The impact of resolution upon entropy and information in coarse-grained models. *The Journal of Chemical Physics* **143,** 243104 (2015).

30. Sanyal, T. & Shell, M. S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *The Journal of Chemical Physics* **145,** 034109 (2016).

31. Dama, J. F., Jin, J. & Voth, G. A. The Theory of Ultra-Coarse-Graining. 3. Coarse-Grained Sites with Rapid Local Equilibrium of Internal States. *Journal of Chemical Theory and Computation* **13,** 1010–1022 (2017).

32. Jin, J. & Voth, G. A. Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems. *Journal of Chemical Theory and Computation* **14,** 2180–2197 (2018).

33. Rosenberger, D. & Van der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Physical Chemistry Chemical Physics* **20,** 6617–6628 (2018).

34. Sanyal, T. & Shell, M. S. Transferable Coarse-Grained Models of Liquid–Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *The Journal of Physical Chemistry B* **122,** 5678–5693 (2018).

35. Izvekov, S. & Voth, G. A. Multiscale coarse graining of liquid-state systems. *The Journal of Chemical Physics* **123,** 134105 (2005).

36. Izvekov, S. & Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *The Journal of Physical Chemistry B* **109,** 2469–2473 (2005).

37. Rudzinski, J. F., Lu, K., Milner, S. T., Maranas, J. K. & Noid, W. G. Extended Ensemble Approach to Transferable Potentials for Low-Resolution Coarse-Grained Models of Ionomers. *Journal of Chemical Theory and Computation* **13,** 2185–2201 (2017).

38. Izvekov, S., Chung, P. W. & Rice, B. M. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *The Journal of Chemical Physics* **133,** 064109 (2010).

39. Moore, J. D. *et al.* A coarse-grain force field for RDX: Density dependent and energy conserving. *The Journal of Chemical Physics* **144,** 104501 (2016).

40. Wagner, J. W., Dannenhoffer-Lafage, T., Jin, J. & Voth, G. A. Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions. *The Journal of Chemical Physics* **147,** 044113 (2017).

41. Dama, J. F. *et al.* The Theory of Ultra-Coarse-Graining. 1. General Principles. *Journal of Chemical Theory and Computation* **9,** 2466–2480 (2013).

42. Davtyan, A., Dama, J. F., Sinitskiy, A. V. & Voth, G. A. The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation. *Journal of Chemical Theory and Computation* **10,** 5265–5275 (2014).

43. Fritz, D., Harmandaris, V. A., Kremer, K. & Van der Vegt, N. F. A. Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules* **42.** 00065, 7579–7588 (2009).

44. Brini, E. & Van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *The Journal of Chemical Physics* **137.** 00009, 154113 (2012).

45. Brini, E., Herbers, C. R., Deichmann, G. & Van der Vegt, N. F. A. Thermodynamic transferability of coarse-grained potentials for polymer–additive systems. *Physical Chemistry Chemical Physics* **14.** 00006, 11896 (2012).

46. Ardham, V. R., Deichmann, G., Van der Vegt, N. F. A. & Leroy, F. Solid-liquid work of adhesion of coarse-grained models of n-hexane on graphene layers derived from the conditional reversible work method. *J. Chem. Phys.* **143,** 243135 (2015).

47. Weingärtner, H. Understanding Ionic Liquids at the Molecular Level: Facts, Problems, and Controversies. *Angewandte Chemie International Edition* **47,** 654–670 (2008).

48. Wang, Y., Jiang, W., Yan, T. & Voth, G. A. Understanding Ionic Liquids through Atomistic and Coarse-Grained Molecular Dynamics Simulations. *Accounts of Chemical Research* **40,** 1193–1199 (2007).

49. Dommert, F., Wendler, K., Berger, R., Delle Site, L. & Holm, C. Force Fields for Studying the Structure and Dynamics of Ionic Liquids: A Critical Review of Recent Developments. *ChemPhysChem* **13,** 1625–1637 (2012).

50. Salanne, M. Simulations of room temperature ionic liquids: from polarizable to coarse-grained force fields. *Physical Chemistry Chemical Physics* **17,** 14270–14279 (2015).

51. Maginn, E. J. Atomistic Simulation of the Thermodynamic and Transport Properties of Ionic Liquids. *Accounts of Chemical Research* **40,** 1200–1207 (2007).

52. Sheridan, Q. R., Schneider, W. F. & Maginn, E. J. Role of Molecular Modeling in the Development of CO_2 –Reactive Ionic Liquids. *Chemical Reviews* **118,** 5242–5260 (2018).

53. Maginn, E. J. Molecular simulation of ionic liquids: current status and future opportunities. *Journal of Physics: Condensed Matter* **21,** 373101 (2009).

54. Zeman, J., Uhlig, F., Smiatek, J. & Holm, C. A coarse-grained polarizable force field for the ionic liquid 1-butyl-3-methylimidazolium hexafluorophosphate. *Journal of Physics: Condensed Matter* **29,** 504004 (2017).

55. Merlet, C., Salanne, M. & Rotenberg, B. New Coarse-Grained Models of Imidazolium Ionic Liquids for Bulk and Interfacial Molecular Simulations. *The Journal of Physical Chemistry C* **116,** 7687–7693 (2012).

56. Wang, Y., Feng, S. & Voth, G. A. Transferable Coarse-Grained Models for Ionic Liquids. *J. Chem. Theory Comput.* **5,** 1091–1098 (2009).

57. Wang, Y., Izvekov, S., Yan, T. & Voth, G. A. Multiscale Coarse-Graining of Ionic Liquids †. *The Journal of Physical Chemistry B* **110,** 3564–3575 (2006).

58. Roy, D. & Maroncelli, M. An Improved Four-Site Ionic Liquid Model. *The Journal of Physical Chemistry B* **114,** 12629–12631 (2010).

59. Arienti, M., Pan, W., Li, X. & Karniadakis, G. Many-body dissipative particle dynamics simulation of liquid/vapor and liquid/solid interactions. *The Journal of Chemical Physics* **134,** 204114 (2011).

60. Jochum, M., Andrienko, D., Kremer, K. & Peter, C. Structure-based coarse-graining in liquid slabs. *The Journal of Chemical Physics* **137,** 064102 (2012).

61. DeLyser, M. R. & Noid, W. G. Extending pressure-matching to inhomogeneous systems via local-density potentials. *The Journal of Chemical Physics* **147,** 134111 (2017).

62. Bhargava, B. L. & Balasubramanian, S. Refined potential model for atomistic simulations of ionic liquid [bmim][PF6]. *The Journal of Chemical Physics* **127,** 114510 (2007).

63. Jorgensen, W. L. & Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field; Application to Potent Anti-HIV Agents. *Journal of chemical theory and computation* **8,** 3895–3801 (2012).

64. Cao, Z. & Voth, G. A. The multiscale coarse-graining method. XI. Accurate interactions based on the centers of charge of coarse-grained sites. *J. Chem. Phys.* **143,** 243116 (2015).

65. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52.** 04126, 255–268 (1984).

66. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

67. Hess, B., Kutzner, C., Van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4,** 435–447 (2008).

68. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117.** 05802, 1–19 (1995).

69. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).

70. Hockney, R. W. & Eastwood, J. W. *Computer simulation using particles.* (A. Hilger, Philadelphia, 1988).

71. Machida, H., Sato, Y. & Smith, R. L. Pressure–volume–temperature (PVT) measurements of ionic liquids ([bmim+][PF6-], [bmim+][BF4-], [bmim+][OcSO4-]) and analysis with the Sanchez–Lacombe equation of state. *Fluid Phase Equilibria* **264,** 147–155 (2008).

72. Machida, H., Taguchi, R., Sato, Y. & Smith Jr., R. L. Measurement and Correlation of High Pressure Densities of Ionic Liquids, 1-Ethyl-3-methylimidazolium l-Lactate ([emim][Lactate]), 2-Hydroxyethyl-trimethylammonium l-Lactate ($[(C_2H_4OH)(CH_3)_3N]$[Lactate]), and 1-Butyl-3-methylimidazolium Chloride ([bmim][Cl]). *Journal of Chemical & Engineering Data* **56,** 923–928 (2011).

73. Ghatee, M. H. & Zolghadr, A. R. Surface tension measurements of imidazolium-based ionic liquids at liquid–vapor equilibrium. *Fluid Phase Equilibria* **263,** 168–175 (2008).

74. Neumann, M. Computer simulation and the dielectric constant at finite wavelength. *Mol. Phys.* **57,** 97–121 (1986).

75. Glättli, A., Daura, X. & Van Gunsteren, W. F. Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. *J. Chem. Phys.* **116,** 9811–9828 (2002).

76. González-Melchor, M., Bresme, F. & Alejandre, J. Molecular dynamics simulations of the surface tension of ionic liquids. *The Journal of Chemical Physics* **122,** 104710 (2005).

77. Hansen, J.-P. & McDonald, I. R. *Theory of simple liquids* 3rd ed (Elsevier / Academic Press, Amsterdam ; Boston, 2007).

78. Tokuda, H., Hayamizu, K., Ishii, K., Susan, M. A. B. H. & Watanabe, M. Physicochemical Properties and Structures of Room Temperature Ionic Liquids. 1. Variation of Anionic Species. *The Journal of Physical Chemistry B* **108,** 16593–16600 (2004).

79. Izvekov, S. & Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *The Journal of Chemical Physics* **125,** 151101 (2006).

80. Hijón, C., Español, P., Vanden-Eijnden, E. & Delgado-Buscalioni, R. Mori–Zwanzig formalism as a practical computational tool. *Faraday Discussions* **144,** 301 (2010).

81. Li, Z., Bian, X., Li, X. & Karniadakis, G. E. Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism. *The Journal of Chemical Physics* **143,** 243128 (2015).

82. Li, Z., Lee, H. S., Darve, E. & Karniadakis, G. E. Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts. *The Journal of Chemical Physics* **146,** 014104 (2017).

83. Van der Vegt, N. F. A. & Deichmann, G. Bottom-up Approach to Represent Dynamic Properties in Coarse-Grained Molecular Simulations. *The Journal of Chemical Physics* **149,** 244114 (2018).

84. Taherian, F., Leroy, F. & Van der Vegt, N. F. A. Interfacial Tension Does Not Drive Asymmetric Nanoscale Electrowetting on Graphene. *Langmuir* **31,** 4686–4695 (2015).

85. Taherian, F., Leroy, F., Heim, L.-O., Bonaccurso, E. & Van der Vegt, N. F. A. Mechanism for Asymmetric Nanoscale Electrowetting of an Ionic Liquid on Graphene. *Langmuir* **32,** 140–150 (2016).

6 Bottom-up Approach to Represent Dynamic Properties in Coarse-Grained Molecular Simulations

Abstract

Several molecular coarse-graining methods have been proposed in recent years to derive chemical- and state-point transferable force fields. While these force fields describe structural and thermodynamic properties in good agreement with fine-grained models and experiments, dynamic properties are usually overestimated. Herein, we examine if long-time dynamic properties of molecular coarse-grained (CG) systems can be correctly represented by employing a dissipative particle dynamics (DPD) thermostat, which is "bottom-up informed" by means of a variant of the Markovian Mori-Zwanzig (MZ) DPD coarse-graining method. We report single-site and multiple-site CG models for a monomer, dimer, and 24mer based on 2,2-dimethyl propane as chemical repeat unit, and report data obtained from MZ-DPD simulations of liquids, polymer solutions, and polymer melts. We find that despite incomplete time scale separation of the molecular CG model, MZ-DPD achieves quantitative accuracy in predicting diffusive dynamics in single-component liquids and polymer solutions (24mers in a dimer solvent). We also find that MZ-DPD simulations of molecular penetrant diffusion in polymer networks do not reach quantitative agreement with the fine-grained model. Modeling diffusion governed by activated barrier crossing of small molecular penetrants in these dense systems requires an accurate description of energy barriers, presumably combined with the treatment of memory effects. The use of a MZ-DPD thermostat extends the scope and applicability of molecular CG models for multicomponent systems where a correct description of the relative diffusion rates of the different components is important.

6.1 Introduction

Coarse-grained (CG) models that link chemistry and properties are important in mesoscale simulations of soft matter. Systematic CG models are derived in a bottom-up approach based on a pre-selected mapping of the degrees of freedom and detailed information of a fine-grained (FG) reference model.[1–3] In this approach, multiple FG sites (e.g. atoms) are usually merged into a single CG interaction site. The bonded and nonbonded interaction potentials used to describe effective interactions between the CG sites can be derived with a variety of approaches.[4–9]

While structural and thermodynamic properties of these models can readily be calculated from conventional molecular dynamics (MD) simulations, the study of dynamic properties is more challenging. The flatter potential energy surface in CG configuration space usually leads an acceleration ("speed-up") of dynamics.[10–14] This speed-up is a useful feature for the purpose of equilibrating complex liquids with long relaxation times. In combination with inverse mapping procedures, it permits studying structural and thermodynamic properties on multiple length scales. It however becomes problematic if dynamic and transport properties are studied. For some problems, scaling factors can be calculated that relate the time scales of the CG and FG representations and allow to predict dynamic properties in the long-time limit (e.g. diffusion) with quantitative accuracy.[15, 16] However, such procedures are not generally applicable, in particular in systems with multiple components where coarse-graining may affect the relative mobilities.[14]

A route to dynamically consistent coarse graining is based on the generalized Langevin equation (GLE). In contrast with Hamiltonian dynamics used in standard MD simulations, this equation of motion includes a friction term that accounts for energy dissipation that occurs in non-elastic collisions of CG beads. In the general case of the GLE, the friction has a time dependence (memory). The dissipation of energy in internal (not explicitly modeled) degrees of freedom of the beads is balanced in the GLE by

a fluctuating force, which is fundamentally related to the friction according to the fluctuation dissipation theorem. The GLE can be derived using the Mori-Zwanzig (MZ) projection operator formalism.[13, 17–20] An important limiting scenario – the Markov limit – corresponds to the situation where the time scale of the fluctuating force is fast compared to the time scale of the momentum of the CG bead. Based on the assumption of time scale separation, and further assumptions discussed later on, one arrives at the dissipative particle dynamics (DPD) equation of motion.[21–24] Applied to molecular CG models, the DPD friction and noise terms account for the fluctuating force associated with molecular vibrations.

In the literature, several studies have been reported on the derivation of MZ-DPD models for systems of Lennard-Jones fluids,[25, 26] repulsive star polymers,[13, 27–29] and liquids.[30–33] Alternative methods of bottom-up derivation of DPD models have been proposed as well.[34] All of the MZ-DPD studies cited above (with the exception of ref. [33]) have in common that the CG mapping is set up such that molecules are represented by only one interaction site in the CG representation. For systems such as star polymers at low density (and other aggressively coarse-grained systems), the Markovian assumption is accurate, hence the velocity autocorrelation function (VACF) of the molecular center of mass is reproduced in agreement with the FG model.[13, 27–29] While this leads to quantitative predictions of transport properties of specific systems, little additional information can be obtained from these models owing to the strong degree of coarse graining. Multiple-bead molecular CG models stay closer to the chemical structure of the molecules, however, with these models the requirement of time scale separation underlying MZ-DPD may not be satisfied as easily. However, as previous studies have shown,[32, 33] long-time diffusion properties may still be accurately represented, despite the lack of complete time scale separation. Based on the initial work by Trément et al.,[33] Lemarchand et al. have, in a recent study, developed molecular CG models for a realistic polymer (*cis*- and *t*rans-polybutadiene) at the experimental density corresponding to 1 atm. ambient pressure.[35] These authors show that the constraint dynamics approach developed by Hijon et al.[13] can be used to derive multiple-bead DPD models. In an earlier work, we have reported a molecular CG hexane model with two sites per molecule.[36] However, the model turned out to not convincingly reproduce the fine-grained dynamics, mainly due to convergence problems of the force correlation functions resulting from constraining few 'slow' degrees of freedom only.

In this work, we propose a method for deriving MZ-DPD molecular CG models using a combination of the effective-force coarse-graining method[8] and the so-called Q-approximation from unconstrained FG trajectories.[20, 27] The advantage of this method over other strategies is that the MZ-DPD force field can be derived from a single FG trajectory without the need of constraining degrees of freedom in the simulation. The method is tested on molecular CG models with one, two, and 24 sites per molecule, each site representing a 2,2-dimethyl propane unit. In addition to MZ-DPD simulations of the pure liquids, we also study the dynamics of polymer solutions (24mer chains in a dimeric solvent) and explore the possibility of modeling transport of small molecules through a 395mer polymer matrix with MZ-DPD. We show that the bottom-up parameterized DPD thermostat successfully repairs the long-time diffusive dynamics in systems (liquids, polymer solutions) where dissipative molecular collisions govern the mobility of the components.

The remainder of the manuscript is structured as follows: In section 6.2, we shortly discuss the fundamental ideas behind the MZ-DPD method and review previous literature on the subject. A description of the specific approach selected for the derivation of the model interactions in this work is provided in sections 6.3 and 6.4. After introducing the model systems and simulation details in section 6.5, we present and discuss our results in 6.6. Section 6.7 contains concluding remarks.

6.2 Mori-Zwanzig DPD

In the Mori-Zwanzig projection operator formalism the Liouville operator of the fine-grained system is split into two parts using the (orthogonal) operators P and Q:[17–19]

$$i\mathscr{L} = Pi\mathscr{L} + Qi\mathscr{L} \tag{6.1}$$

$$Q = \mathscr{I} - P. \tag{6.2}$$

The operator P denotes a projection onto the N^P degrees of freedom (DoF) remaining in the CG representation, $\mathscr{I}$ denotes the identity operator. In a sensibly chosen CG representation, these DoF, which we will refer to as P-DoF, relax relatively slow. The remaining DoF (Q-DoF) relax fast and are not explicitly represented in the CG model but will be implicitly modeled by dissipative and noise terms.

The coarse-grained equation of motion (EOM) for the N^P momentum vector $\mathbf{p}^P$ emerging from this projection of the dynamics is the GLE, which contains conservative ($\mathbf{F}^{\mathrm{C}}$), dissipative ($\mathbf{F}^{\mathrm{D}}$) and random forces ($\mathbf{F}^{\mathrm{R}}$):[13, 20, 37]

$$\frac{d\mathbf{p}^P}{dt}(t) = \mathbf{F}^{\mathrm{C}} + \mathbf{F}^{\mathrm{D}} + \mathbf{F}^{\mathrm{R}} \tag{6.3}$$

$$\frac{d\mathbf{p}^P}{dt}(t) = -\frac{\partial}{\partial \mathbf{r}^P} U^{\mathrm{CG}}(\mathbf{r}^P(t)) - \int_0^t \mathbf{\Gamma}(t-t', \mathbf{r}^P)\mathbf{v}^P(t')\, dt' + \delta\mathbf{F}^Q. \tag{6.4}$$

The GLE is non-Markovian (i.e. it incorporates explicit memory effects), because the momentum evolution does not only depend on the $3N^P$ CG coordinates $\mathbf{r}^P$ and velocities $\mathbf{v}^P$ at the current time t but also on all earlier times. This is expressed by the integral in the second term in which $\mathbf{\Gamma}(t)$ is a $(3N^P \times 3N^P)$ memory kernel of the fluctuating force $\delta\mathbf{F}^Q$. The last two terms of the GLE are related through the fluctuation-dissipation theorem, so the knowledge of the multi-body potential of mean force U^{CG} associated with the CG mapping and the memory kernel $\mathbf{\Gamma}(t)$ is, in principle, sufficient to construct a CG model. To be useful in practise, U^{CG} is usually assumed to be pairwise additive. The time evolution of the momentum of a single site i reads:

$$\frac{d\mathbf{p}_i^P}{dt} = -\frac{\partial}{\partial \mathbf{r}_i^P} U^{\mathrm{CG}}(\mathbf{r}^P(t)) - \sum_j \int_0^t \mathbf{\Gamma}_{ij}(t-t', \mathbf{r}^P)\mathbf{v}_j^P(t')\, dt' + \delta\mathbf{F}_i^Q, \tag{6.5}$$

where $\mathbf{\Gamma}_{ij}$ denotes the 3×3 submatrix of $\mathbf{\Gamma}$ for the pair of sites i and j. Newton's third law requires:

$$\mathbf{\Gamma}_{ii} = -\sum_{j \neq i} \mathbf{\Gamma}_{ij}. \tag{6.6}$$

Therefore, we can reformulate the EOM (Eq. 6.5) using relative velocities $\mathbf{v}_{ij}^P = \mathbf{v}_i^P - \mathbf{v}_j^P$:

$$\frac{d\mathbf{p}_i^P}{dt} = -\frac{\partial}{\partial \mathbf{r}_i^P} U^{\mathrm{CG}}(\mathbf{r}^P(t)) - \sum_{j \neq i} \int_0^t \mathbf{\Gamma}_{ij}(t-t', \mathbf{r}^P)\mathbf{v}_{ij}^P(t')\, dt' + \delta\mathbf{F}_i^Q. \tag{6.7}$$

CG models that use the GLE with explicit memory terms have been derived and studied in the past.[28, 38, 39] These simulations, however, are computationally more expensive. The approach commonly chosen is a further simplification of the CG EOM to obtain the DPD-like MZ-DPD EOM for CG site i:[13, 20, 37]

$$\frac{d\mathbf{p}_i^P}{dt} = \sum_{j \neq i} \left[-\frac{\partial}{\partial \mathbf{r}_i^P} U_{ij}^{\mathrm{CG}}(r_{ij}^P) - \gamma_{ij}(r_{ij}^P)\mathbf{v}_{ij}^P + \sigma_{ij}(r_{ij}^P)\xi_{ij} \right], \tag{6.8}$$

where γ_{ij} denotes a DPD pair friction and ξ_{ij} is a vector of Gaussian random variables with zero mean and unit variance. The strength of the random force $\sigma_{ij} = \sqrt{2k_B T\gamma_{ij}}$ depends on γ_{ij} through the fluctuation dissipation theorem.[22]

The transformation of the GLE (Eq. 6.4) into the MZ-DPD EOM (Eq. 6.8) relies on two assumptions about the nature of the coarse-grained system. Firstly, we assume that the multi-body potential of mean force U^{CG} can be approximated by a combination of 2-, 3-, and 4-body bonded (intramolecular) potentials describing the molecular geometry and nonbonded pair potentials. This assumption is also used in conventional bottom-up coarse-graining methods and is well justified for many soft matter systems.[2] Secondly, we assume complete separation of time scales of the two parts of the dynamics: we assume that the (fast) Q-DoF relax infinitely fast on the relaxation time scale of the (slow) P-DoF. The second assumption allows us to replace the memory kernel in Eq. (6.7) with a Dirac delta function $2\gamma_{ij}\delta(t-t')$, which, upon integration, yields a friction matrix γ_{ij}. This assumption is well-justified for systems with low density and/or a high degree of coarse-graining (a large number of FG sites per CG site).[13, 26, 27, 29] For dense systems described with molecular coarse-grained models, this assumption is however not strictly valid. Because in these models the masses of the CG and FG particles do not differ significantly and high-frequency particle collisions lead to a relatively fast decorrelation of CG particle velocities, the relaxation of the P- and Q-DoF are comparable.[32, 36]

6.2.1 Systems Studied with MZ-DPD

Several systematic bottom-up MZ-CG models for molecular systems have been developed for modeling star polymer systems with a mapping chosen such that all atoms of the fine-grained molecules are mapped onto a single CG interaction site.[13, 26, 28, 29, 39] In other studies, fluids of small molecules have been considered as model systems and the coarse-graining is performed by representing a cluster of several molecules by a single CG interaction site.[25, 27, 30–32] The results obtained in these studies illustrate the physical principles behind the projection operator formalism and highlight the importance of time scale separation for the exact validity of the Markovian assumption in the derivation of the DPD EOM. It can be concluded that whenever the time scales of P- and Q-DoF are clearly separated, a MZ-DPD model is sufficient for the description of both the dynamics on short and long time scales. This is usually the case when the density is low or the degree of coarse-graining is high. In these cases, the CG site velocities (which can be regarded as the fastest of the supposedly slow P-DoF) decorrelate on time scales that are long compared to the time scale of the memory kernel.

In studies that instead require the use of molecular CG models, the density at ambient pressure is usually high and the models show, by definition, a small to medium degree of coarse-graining. These features cannot be fundamentally changed because the high density is not chosen by the modeler but rather imposed by the intermolecular interactions. At high densities, the decorrelation of CG site velocities is faster since collisions between molecules become more frequent. It is for this reason that the time scales of retained and omitted DoF do not strictly separate in simulations of moderately coarse-grained systems at realistic liquid-state densities. The validity of the Markovian assumption is critical to the success of coarse-graining with completely consistent dynamics.[26, 32, 33, 36, 40] Short time dynamics in molecular CG systems with high particle number densities is therefore usually not accurately reproduced, although in many cases the self-diffusion coefficients can be reproduced to a satisfying degree.[25, 30–33, 35, 36, 40–42]

As discussed before, we aim to investigate the possibility of deriving MZ-DPD models using CG mappings with more than one site per molecule which are based on systematic coarse-graining methods. To our knowledge, only two studies have so far been published on MZ-DPD models for molecular CG mappings, apart from our own earlier work (ref. [36]). In their study on linear alkanes, Trément et al. model a decane molecule as a dimer consisting of two pentane molecules.[33] Although the model was derived using FG simulations of the pentane monomer, the resulting model was capable of reproducing self-diffusion rates and viscosities of liquid decane to a satisfying degree. In a following study, a

polybutadiene molecule was coarse-grained using a number molecular mappings with increasing degree of coarse-graining.[35] The resulting MZ-DPD models are able to reproduce the self-diffusion rate and viscosity. The quality of the models increases with increasing degree of coarse-graining, a result that is expected by theory.

6.3 Coarse-Graining with a Molecular Mapping

The projection operator formalism describes a theoretical framework for the calculation of the three forces in the GLE (Eq. 6.4). While the GLE is exact, additional assumptions are required to obtain a useful equation of motion that can be solved with limited computational resources.

6.3.1 The Conservative Interaction

The conservative force depends on the positions of the degrees of freedom (CG interaction sites) that are explicitly modeled in the CG model (P-DoF) $\mathbf{r}^P$. It can be written as gradient of the potential of mean force U^{CG}:[13]

$$\mathbf{F}_i^{\mathrm{C}} = -\frac{\partial}{\partial \mathbf{r}_i^P} U^{\mathrm{CG}}(\mathbf{r}^P). \tag{6.9}$$

This expression is exact but not useful in computer simulations. The PMF depends in a multi-body manner on the positions of *all* CG interaction sites. The exact evaluation of the conservative force is therefore a problem of huge dimension that cannot be undertaken with finite computational resources. When constructing a bottom-up CG model, an approximation has to be found that only includes low orders of correlations between positions, ideally reducing the problem to a sum of pairwise forces.

The most simple approximation, which is used in many applications of MZ-DPD models,[13, 27, 33, 35] is calculating a pair-PMF from the radial distribution function $g_{ij}(r)$ of the CG sites of interest calculated from a FG trajectory:

$$U_{ij}^{\mathrm{pair}}(r) = -k_{\mathrm{B}}T \ln\left[g_{ij}(r)\right]. \tag{6.10}$$

While this approximation is representable for the case of very coarse models at moderate densities, it usually falls short in reproducing several important properties, such as liquid structure, pressure or density in molecular CG models of soft matter at higher density.[3]

Various bottom-up coarse graining algorithms to derive pair potentials for molecular CG models have been formulated in the literature. In these methods, structure-matching, force-matching, or coupling free energies are used to derive pair potentials which give a more realistic model of the multi-body potential of mean force than the use of pair PMFs.[2] Some of these advanced coarse-graining methods have already been applied in the context of deriving MZ-DPD models.[32, 36, 42]

In this work, we will use the effective force coarse-graining (EF-CG) method first published by Wang et al. in 2009.[8] In this method, the CG interaction forces are obtained directly from the FG pair forces. It is straightforward to implement and the method allows to derive the potentials directly from a FG simulation trajectory. The models derived with this method show good representability in NVT and NpT simulations and are (to a certain degree) temperature transferable. The EF-CG method is similar in spirit to the approach of Fritz et al. for polymer coarse-graining.[15]

Because of its formulation in terms of pairwise forces, the EF-CG method naturally lends itself in developing MZ-DPD models together with the pairwise friction calculation introduced on the basis of the work of Kinjo and Hyodo.[20, 26–28]

6.3.2 The Dissipative Interaction

Let us reiterate the definition of the dissipative force in the GLE emerging from the projection operator formalism:

$$\mathbf{F}^{\mathrm{D}}(t) = -\int_0^t \mathbf{\Gamma}(t-t', \mathbf{r}^P)\mathbf{v}(t')\, dt'. \tag{6.11}$$

The memory kernel $\mathbf{\Gamma}(t)$ is a $3N^{\mathrm{CG}} \times 3N^{\mathrm{CG}}$-matrix where N^{CG} is the number of sites in the CG model. The ij-th submatrix of this matrix (representing the dissipative interaction between sites i and j) can be calculated by use of the fluctuation dissipation theorem from the fluctuating force:[13]

$$\mathbf{\Gamma}_{ij}(t) = \frac{1}{k_{\mathrm{B}}T}\left\langle \delta\mathbf{F}_i^Q(0) \otimes \delta\mathbf{F}_j^Q(t)\right\rangle \tag{6.12}$$

$$= \frac{1}{k_{\mathrm{B}}T}\left\langle \delta\mathbf{F}_i^Q(0) \otimes e^{QiLt}\delta\mathbf{F}_j^Q(0)\right\rangle, \tag{6.13}$$

where $\delta\mathbf{F}_i^Q$ is the fluctuating force on site i. It is important to note that this equality is only exact if the time correlation function of the fluctuating force is calculated from the time evolution determined by the propagator in the projected Q space. This dynamics, unfortunately, cannot be simulated directly.[13]

Several ideas have been brought forward to address this issue and enable an approximate calculation of the friction matrix. Hijon et al. have proposed a simulation method that extracts the dynamics of the Q-DoF by constraining the 'slow' P-DoF to their initial positions.[13] In this case, the resulting constraint dynamics is used as approximation to the Q-dynamics and pair friction functions can be calculated from the trajectory sampled with the modified dynamics. This method has been applied to derive models for organic liquids and polymers.[33, 35, 36] Challenges may arise when using the constraint dynamics approach for systems where the time scales of P- and Q-DoF are close.[35, 36] Lemarchand et al. have found that constraining the 'slow' DoF may also alter the dynamics of 'fast' DoF in polymer systems.[35] They have identified additional Q-DoF that become 'slow' upon the introduction of constraints to the dynamics of the systems. They further assumed that the slowly decaying contribution to the force fluctuation correlation function does not add in a significant manner to the DPD friction. The resulting DPD model was successfully used to simulate the diffusion of polymer molecules in a melt.

Another interesting approach for the approximation of the Q-dynamics has recently been proposed and applied by Izvekov.[42, 43] Here the Q-dynamics is approximated by a momentum-dependent Hermite series of second order.

In the most simple approximation we assume that the real dynamics is equal to the Q-dynamics (Q-approximation):

$$e^{QiLt}\delta\mathbf{F}_j^Q(0) \approx e^{iLt}\delta\mathbf{F}_j^Q(0) \approx \mathbf{F}_j(t) - \mathbf{F}_j^{\mathrm{C}}, \tag{6.14}$$

where $\mathbf{F}_j(t)$ is the instantaneous force on site j and $\mathbf{F}_j^{\mathrm{C}}$ the mean force. This approach has its caveats, among which is the so-called plateau problem, which have been discussed in detail elsewhere.[13, 32, 42] Nonetheless it has been used in various studies and is tempting because of its ease of implementation.[26–29, 32]

Starting from the Q-approximation and the above-mentioned assumptions of pairwise additivity and Markovian behavior, Kinjo and Hyodo[20] derived a scheme for calculating MZ-DPD dissipative forces which is formulated in terms of pair interactions and is used in this work.

6.4 Derivation of a MZ-DPD model

In the EF-CG method, the conservative force between two CG sites is calculated from the sum of the atomistic pair forces between atoms a and b ($\mathbf{F}_{ab}$) belonging to sites i and j respectively:

$$\mathbf{F}_{ij}^{\mathrm{C}}(r) = \left\langle \sum_{a\in i}\sum_{b\in j} \mathbf{F}_{ab}\cdot\mathbf{e}_{ij} \right\rangle_{r_{ij}=r} \mathbf{e}_{ij}, \tag{6.15}$$

where $\mathbf{e}_{ij}$ denotes the unit vector parallel to the i-j distance vector. The angular brackets $\langle ... \rangle_{r_{ij}=r}$ denote averaging over all pairs with $r_{ij} = r$ in an unconstrained FG trajectory.

The DPD pair friction in the MZ-DPD model is expressed through two components that are parallel ($\gamma_{ij}^{\parallel}$) and perpendicular ($\gamma_{ij}^{\perp}$) to $\mathbf{r}_{ij}$, respectively:

$$\gamma_{ij}\mathbf{v}_{ij} = \left(\gamma_{ij}^{\parallel}(\mathbf{e}_{ij}\otimes\mathbf{e}_{ij}) + \gamma_{ij}^{\perp}(\mathscr{I} - \mathbf{e}_{ij}\otimes\mathbf{e}_{ij})\right)\mathbf{v}_{ij}. \tag{6.16}$$

In the Q-approximation, these pair friction functions can be calculated from the fluctuating force, which is approximated by subtracting the conservative force of the CG model from the instantaneous force in a FG trajectory:

$$k_{\mathrm{B}}T\,\gamma_{ij}^{\parallel/\perp}(t,r) = \left\langle \left(\delta\mathbf{F}_{ij}^{Q}(0)\cdot\mathbf{e}_{ij}^{\parallel/\perp}\right)\left(\delta\mathbf{F}_{ij}^{Q}(t)\cdot\mathbf{e}_{ij}^{\parallel/\perp}\right)\right\rangle_{r_{ij}=r} \tag{6.17}$$

$$\approx \left\langle \left(\Delta\mathbf{F}_{ij}(0)\cdot\mathbf{e}_{ij}^{\parallel/\perp}\right)\left(\Delta\mathbf{F}_{ij}(t)\cdot\mathbf{e}_{ij}^{\parallel/\perp}\right)\right\rangle_{r_{ij}(t)=r}, \tag{6.18}$$

where $\Delta\mathbf{F}_{ij}(t) = \mathbf{F}_{ij}(t) - \mathbf{F}_{ij}^{\mathrm{C}}(r_{ij}(t))$, $\mathbf{e}_{ij}^{\parallel} = \mathbf{e}_{ij}$, and $\mathbf{e}_{ij}^{\perp}$ is a unit vector chosen such that $\mathbf{e}_{ij}^{\parallel}\cdot\mathbf{e}_{ij}^{\perp} = 0$.

In order to obtain Eq. (6.17) from Eq. (6.12), it is assumed that the force correlation function in Eq. (6.12) is dominated by the contribution of the direct pair interactions ($\delta\mathbf{F}_{ij}^{Q}$), and that multi-body contributions vanish, i.e. $\left\langle \delta\mathbf{F}_{ij}^{Q}\otimes\delta\mathbf{F}_{ik}^{Q}\right\rangle_{j\neq k} = 0$. This leaves us with a DPD pair friction, which can be calculated from direct (nonbonded) pair interactions, an approach that can be used well in conjunction with the EF-CG method, which is also based exclusively on pair interactions in the FG system.

Using the Markovian assumption, we obtain the time-independent DPD pair friction function through time integration of the correlation function:

$$\gamma_{ij}^{\parallel/\perp}(r) = \frac{1}{k_{\mathrm{B}}T}\int_0^{\infty}\gamma_{ij}^{\parallel/\perp}(\tau,r)\,d\tau. \tag{6.19}$$

One of the caveats of the Q-approximation is that this integral does not converge to a finite plateau value but converges to zero on long time scales (plateau problem).[13, 32, 42] However, it is still possible to derive a MZ-DPD model, if we assume convergence after a short time, before the onset of the decay of the integral value, effectively extracting the short time part of the decay of the force correlation function.[27, 32] A detailed discussion of the procedure used to calculate the pair-friction functions in this work is given in section 6.6.

6.5 Computational Details

6.5.1 Model systems

In an earlier work, we successfully developed a MZ-DPD force field for neopentane, based on conditional reversible work potentials and constraint dynamics simulations in the liquid.[36] In this work, we derive

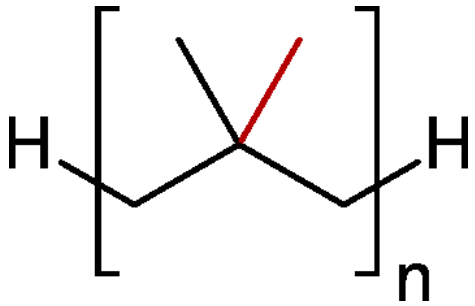


Figure 6.1: Chemical structure of the simulated poly-neopentane. CG interaction sites are located at the center of mass of each repeating unit. The orientation of the red bond is used to analyze the relaxation dynamics of the internal DoF within a site.

another MZ-DPD model for neopentane using EF-CG potentials and the Q-approximation for the pair friction calculation. We then generate more complex test systems by connecting multiple neopentane molecules to form a polyneopentane (poly-NEP) chain (Fig. 6.1). The mapping scheme is always chosen such that one neopentane unit is represented by one site in the CG model.

All systems are simulated at constant NVT using three different simulation methods:

1. Fine-grained molecular dynamics (FG-MD) with the united atom representation of the TraPPE-UA model[44]

2. Coarse-grained molecular dynamics (CG-MD) using the Newtonian EOM

3. Mori-Zwanzig DPD of the coarse-grained system (MZ-DPD)

FG-MD and CG-MD simulations are performed using Gromacs 4.6.7,[45] MZ-DPD simulations are performed using LAMMPS.[46]

Fine-Grained Force Field

The TraPPE-UA force field is used as FG model.[44] The nonbonded interactions are modeled using Lennard-Jones potentials with a cutoff distance of 1.4 nm and tail corrections for pressure and energy. No charges are present in this model. The temperature is kept constant by means of a Nosé-Hoover thermostat with a time constant of 1 ps.[47] This thermostat is also used for the CG-MD simulations. In preparation of the productive NVT simulations, NpT simulations in FG resolution are performed to determine the average mass density. The Parrinello-Rahman barostat is used in all these simulations with a time constant of 5 ps.[48] The time step in all simulations is 1 fs. More details on the force fields used in the CG simulations are presented in the following section.

Pure Liquids

We study liquid phases of 500 neopentane molecules and 250 poly-NEP dimer molecules under NVT conditions at a temperature of 275 K. The edge lengths of the cubic box are 4.6 nm (monomer) and 4.3 nm (dimer) corresponding to mass densities of 606 kg m^{-3} (monomer) and 745 kg m^{-3} (dimer), the mean density in a FG NpT simulation at a pressure of 1 bar. Further, a system of 100 24mers is studied at 500 K in a cubic box with an edge length of 7 nm (mass density 807 kg m^{-3}). The simulation lengths are 10 ns for the small molecules and 500 ns for the 24mer system.

Polymer-Solvent Mixtures

In addition to the pure systems, we also study mixtures of dimers and 24mers: one system (mixture P75S25) contains 75 24mers, corresponding to 1800 polymer (P) beads, and 300 dimers, corresponding to 600 solvent (S) beads. The other system (mixture P25S75) contains 25 24mers, corresponding to 600 P beads, and 900 dimers, corresponding to 1800 S beads. Both systems are simulated under NVT conditions at 500 K (P75S25) and 400 K (P25S75) in cubic boxes of edge length 7.39 nm (P75S25) and 7.15 nm (P25S75), corresponding to the respective mean densities of the FG model at 1 bar pressure (P75S25: 695 kg m^{-3}; P25S75: 680 kg m^{-3}). The simulation length is 100 ns.

Polymer Matrix with Penetrant Molecules

A system is set up for the study of small-molecule penetrant diffusion in a poly-NEP polymer network. To this end, we place five 395mers, ten dimers and ten monomers in a cubic box with edge length 6.75 nm (773 kg m^{-3}) at a temperature of 500 K. Starting from an equilibrated structure, we simulate 100 ns with each model. We perform a second simulation where we restrain the motion of the polymer matrix to mimic the dynamics of a glassy network at the given temperature. To this end, we apply a harmonic potential to the sites on the backbone of the polymer:

$$U_{\text{restrain}} = k_{\text{restrain}}(\mathbf{r}-\mathbf{r}_0)^2, \tag{6.20}$$

where $k_{\text{restrain}} = 50\,\text{kJ}\,\text{mol}^{-1}\,\text{nm}^{-1}$, $\mathbf{r}$ is the position of a site and $\mathbf{r}_0$ is its initial position. In the FG-MD simulations this restraint only acts on the quarternary carbon in the backbone.

6.6 Results and Discussion

6.6.1 CG Models

Potentials

The non-bonded pair potentials used for the CG simulations of the pure systems are shown in figure 6.2. The attractive tails decay monotonically as typically observed for pair potentials of CG models calculated using coupling-free energy methods.[8, 9] All three potentials show similar characteristic features: the potentials have a minimum of ca. 3 kJ mol^{-1} at a distance of ca. 0.65 nm. The non-bonded pair potential derived for interaction sites (repeat units) on the 24mer is the least attractive. This is not surprising, since it was derived at a higher temperature than the other potentials, which enhances the contribution from the (repulsive) entropic part of the free energy of interaction. The same pair potential is used to describe intramolecular nonbonded interactions between all sites within the 24mer, except for first, second, and third neighbor sites whose interactions are modeled with bonded interaction potentials. In all CG simulations of polymer-solvent mixtures and polymer matrices with penetrant molecules discussed below, the nonbonded interactions between all components in the system are described with the same pair potential, which corresponds with the one derived for the 24mer sites.

The bonded interactions are obtained by Boltzmann-inverting distributions of bond lengths, bending angles, and torsional angles, which are gathered from a 100 ns simulation of one molecule in vacuo. The bond stretching potential is fitted by a harmonic potential:

$$U_{\text{bond}}(r) = k_{\text{bond}}(r-r_0)^2. \tag{6.21}$$

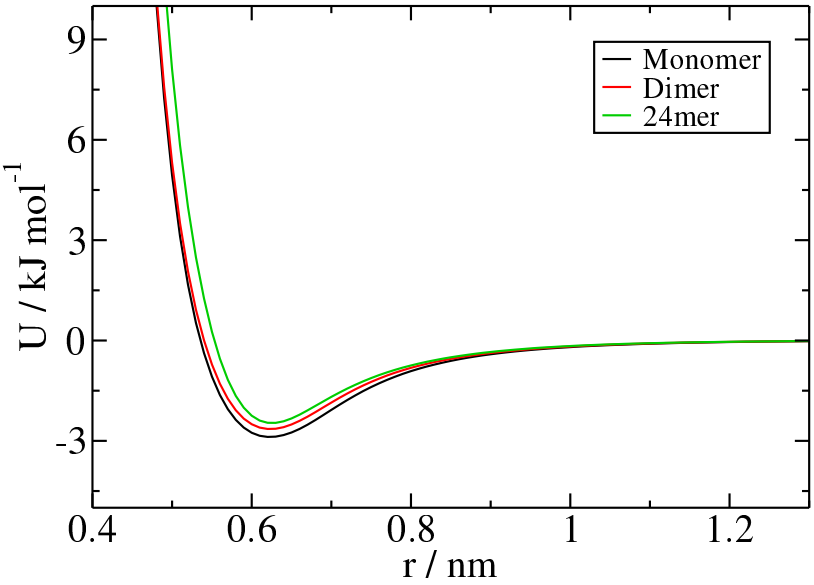


Figure 6.2: Nonbonded EF-CG pair potentials used for the simulations of the pure component systems. The pair potentials for the monomer and dimer were derived at 275 K; for the 24mer, the pair potential was derived at 500 K.

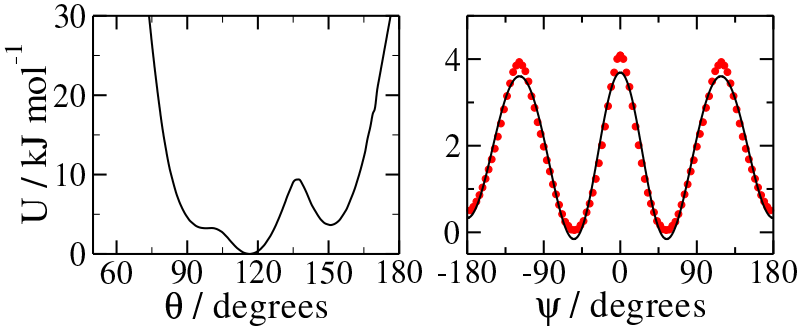


Figure 6.3: Bond angle (left) and torsion (right) potential of the poly-NEP molecule. The red dots in the right panel were obtained from Boltzmann inversion. The black line corresponds to the fit (Eq. 6.22).

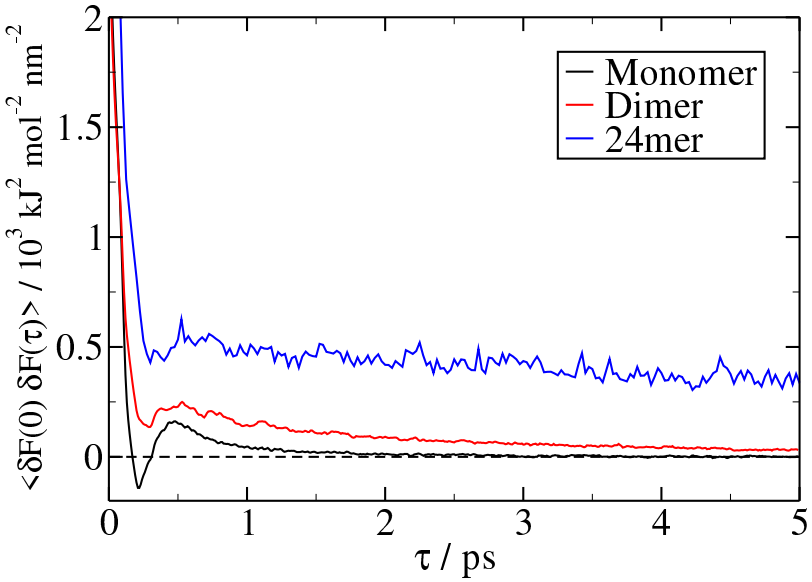


Figure 6.4: Force autocorrelation functions (FACFs) at an inter-site distance of 0.6 nm. For the monomer liquid, the FACF converges to zero quickly. For longer chains, the decay becomes very slow on longer time scales.

The dimer bond stretching is modeled with $k_{\text{bond}} = 24075$ kJ mol^{-1} nm^{-2} and $r_0 = 0.404$ nm, while for the 24mer we use $k_{\text{bond}} = 23250$ kJ mol^{-1} nm^{-2} and $r_0 = 0.396$ nm. For the bond angle bending potential a tabulated potential is used in the CG simulations (Fig. 6.3). The torsional potential of the polymer is modeled using a sum of 6 cosine functions of increasing power:

$$U_{\text{tors}} = \sum_{i=0}^{5} c_i (\cos\psi)^i, \tag{6.22}$$

where $c_0 = 2.18$ kJ mol^{-1}, $c_1 =$ -5.02 kJ mol^{-1}, $c_2 =$ -2.17 kJ mol^{-1}, $c_3 = 4.94$ kJ mol^{-1}, $c_4 = 2.00$ kJ mol^{-1}, $c_5 = 1.76$ kJ mol^{-1}.

DPD Pair Frictions

The force autocorrelation functions (FACFs, Eq. (6.18)) are calculated from fine-grained NVT simulations. The FACFs are shown in figure 6.4. To satisfy the Markovian assumption, the FACFs must decay to zero rapidly in comparison with the decay of the VACF of the CG sites (i.e. the 'slow' DoF of the system). We find, however, that the FACFs have a long time tail. After a fast initial decay, the FACF decays on a time scale that is larger than the time scale of the decay of the VACF. The long time decay is least significant for the monomer and becomes more important in the dimer and 24mer systems. This behavior is clearly undesired since the Markovian assumption made in the derivation of the MZ-DPD simulation algorithm requires a complete relaxation of the Q-DoF on small time scales.

One reason for this type of behavior may be related to modeling errors. Izvekov describes similar observations for a system of liquid nitromethane. [42] More specifically, slowly decaying orientations of the multi-site molecules may cause the average fluctuating force to be non-zero on short time scales, because the average force acting between two groups of atoms on the time scale of a few ps (instantaneous mean force) will probably not be equal to the mean force used as conservative interaction in the

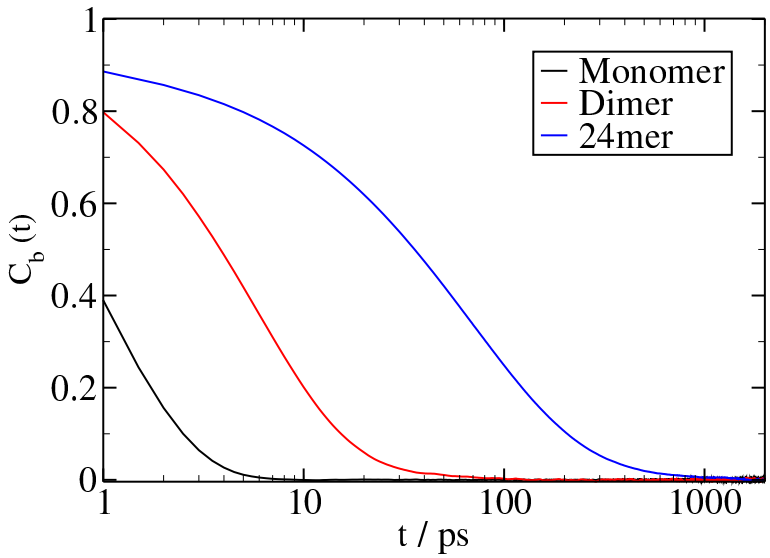


Figure 6.5: Time correlation function $C_b(t)$ defined in Eq. (6.23) describing the dynamics of rotation of a methyl group in the poly-NEP backbone.

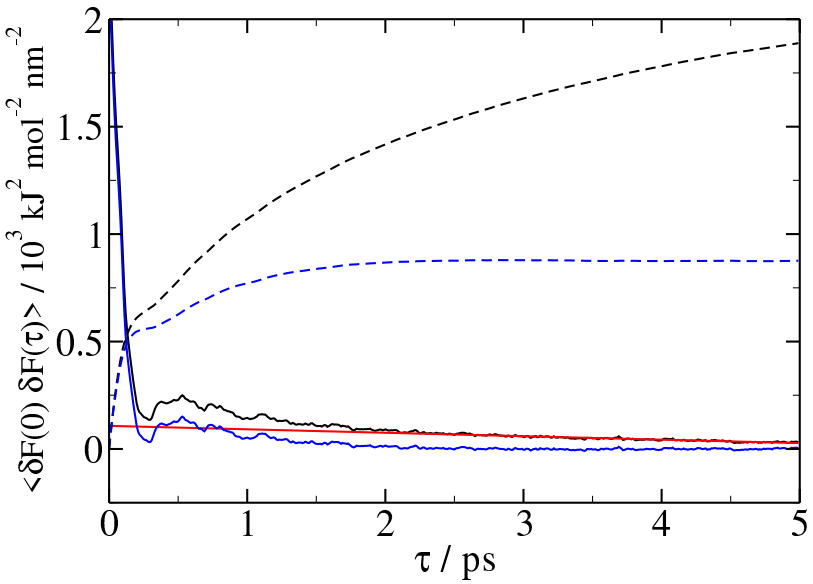


Figure 6.6: Illustration of the fitting procedure employed to extract the short time component of the FACF. The black curve is the FACF calculated from the FG dimer trajectory for the intermolecular site-site distance of 0.6 nm. Its slow time decay on long time scales is fitted with a linear function (red line). The blue curve, which is obtained by subtracting the linear function from the original data, converges faster to zero. The dashed lines are the integrals of their respective solid counter part (not to scale).

CG model. This mismatch in the instantaneous and theoretical mean forces will decay with increasing time, but especially when there are slowly relaxing orientations involved, the time scales of this decay may be rather large.

Figure 6.5 shows the time correlation function C_b which describes the rotation dynamics of a methyl group in the poly-NEP backbone:

$$C_\mathrm{b}(t) = \langle \mathbf{e}_\mathrm{b}(0) \cdot \mathbf{e}_\mathrm{b}(t) \rangle \,, \tag{6.23}$$

where $\mathbf{e}_\mathrm{b}$ is the unit vector parallel to the bond connecting a methyl group to the quarternary carbon of the backbone (the bond drawn red in figure 6.1). Ideally, $C_\mathrm{b}(t)$ should relax fast in comparison to the P-DoF of the system, since it describes a DoF no longer present in the CG model. However, we find that the relaxation time is only relatively small for the monomer molecule and increases with increasing chain length. We fit C_b using an exponential function $C_\mathrm{b} = \exp(-t/\tau_\mathrm{b})$ and find that τ_b is ca. 1 ps for the monomer, 6 ps for the dimer, and 70 ps for the 24mer. This finding is in agreement with the observed long time relaxation in the FACFs (Fig. 6.4). Calculating the FACFs from a simulation where the 'slow' DoF are constrained does not resolve the issue of long time tails; it leads to an even slower relaxation of orientations because these relaxation processes are coupled to the relaxation of the P-DoF of the CG model.[35, 36]

In order to extract the short time component of the FACF, we fit to the long time decay (in a range of 2.5 ps to 5 ps) with a linear fitting function. This linear approximation of the long time behavior is then subtracted from the FACF and the integral over this short time contribution is used to calculate the DPD pair friction function. This integral now shows the desired convergence to a plateau value on comparatively short time scales. An illustration of this procedure is shown in figure 6.6 using the FACF of the dimer sites at $r_{ij} = 0.6$ nm. The choice of a linear fitting function is not motivated from the physical form of the long time tail (which is in fact non-linear). We choose the linear fitting function as an approximation, because it closely approximates the decay on long time scales and because it is a robust function in the automated numerical implementation of the friction calculation.

Using this procedure, we calculate pair friction functions for all pure systems and mixtures. Figures 6.7 and 6.8 show the friction functions obtained for pure component liquids and polymer-solvent mixtures, respectively. The pair friction for the 24mer is calculated by averaging over all sites in the chain. All pair friction functions are largest at short distance and become approximately zero at a distance close to the minimum of the nonbonded pair potentials (Fig. 6.2). This nicely illustrates the physical significance of the DPD pair friction: it effectively accounts for the excitation of the Q-DoF in non-elastic particle collisions. The transversal DPD frictions are significantly smaller than the parallel ones, a finding which is well in line with earlier studies on the subject of MZ-DPD models.

For the use in MZ-DPD simulations, pair friction tables are precompiled using spline interpolation between calculated friction values and linear extrapolation for values below the lowest sampled distance with a distance increment of 0.001 nm. For the simulations of polymer matrices with penetrant molecules, we decided to use the pair frictions parametrized for mixture P75S25 such that all sites representing small molecules (monomer and dimer) are represented by dimer-sites. This choice is motivated by the realization that the pair friction does not depend strongly on the temperature and composition of the system.

6.6.2 Pure Components

Structural Properties

To evaluate the nonbonded part of the CG model, we compare the radial distribution functions (RDFs) from the FG- and CG-MD simulations (Fig. 6.9). We observe that the short-range, excluded volume, region and the first maximum of the RDFs is well reproduced in all simulations. For liquid neopentane

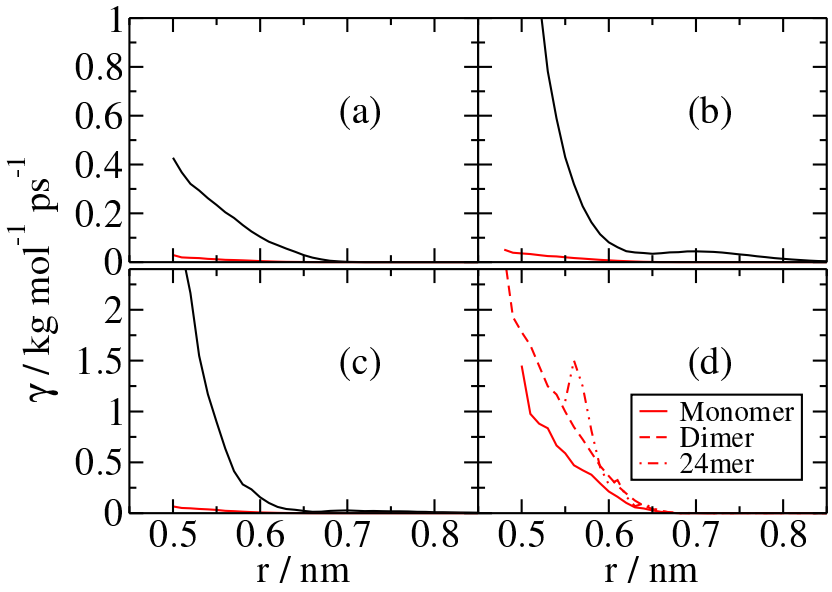


Figure 6.7: DPD pair friction functions used in the simulation of pure liquid systems of the monomer (a), the dimer (b), and the 24mer (c). Black lines denote the parallel friction component, red lines denote the perpendicular component. Panel (d) shows the three transversal pair frictions multiplied by a factor of 50.

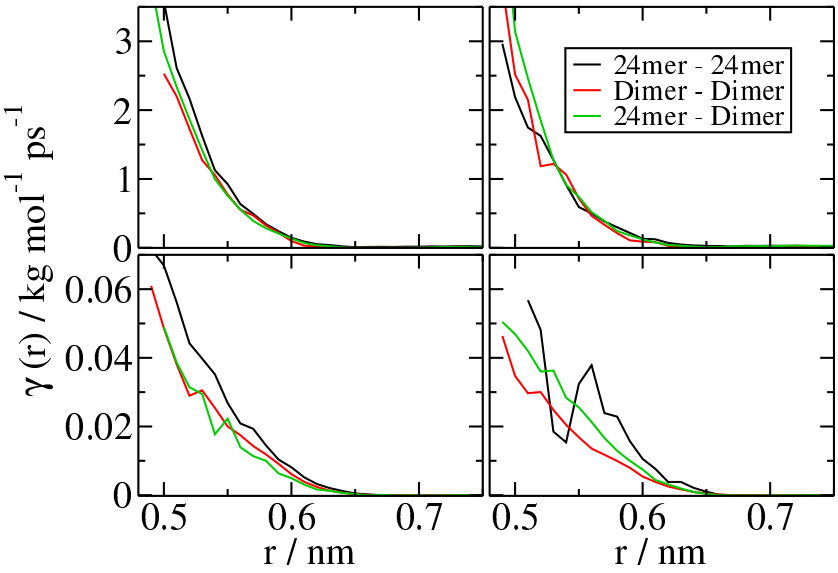


Figure 6.8: DPD pair friction functions calculated from FG simulations of polymer-solvent mixture P75S25 (left) and mixture P25S75 (right). The upper panels and lower panels show the parallel and transversal components of the DPD pair friction, respectively.

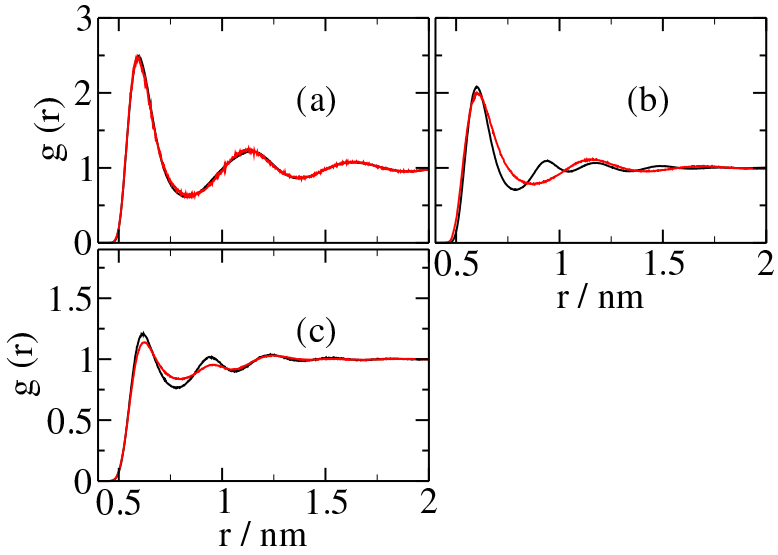


Figure 6.9: Site-site radial distribution function $g(r)$ calculated from simulations of the monomer (a), dimer (b), and 24mer (c) in the liquid phase using FG-MD (black lines) and CG-MD (red lines) models. The RDFs obtained from MZ-DPD simulations are identical to those obtained from CG-MD simulations.

(the monomer), also the remaining part of the RDF is reproduced almost to a quantitative degree. Qualitative disagreements are observed with the dimer model: while the FG model shows two consecutive maxima around 1 nm, the CG model only predicts one, broader, maximum at the distance of the third maximum of the FG model. For the 24mer, the essential features of the RDF are reproduced by the CG model but the amplitudes of the minima and maxima are underestimated by the CG model. In addition to the RDF, we calculate the characteristic ratio C_n to compare the internal structure of the polymer in the FG and CG representations:

$$C_n = \frac{\langle r_n^2 \rangle}{nb^2}. \tag{6.24}$$

In Eq. (6.24), $\langle r_n^2 \rangle$ is the mean-squared distance between two repeat units separated by n bonds and $b = 0.404$ nm is the mean bond length between two repeat units. Figure 6.10 shows that C_n of the FG-MD representation is reproduced well with both CG models. Further, we calculate the radius of gyration of the 24mer. The mean radius of gyration obtained with the FG-MD model is 1.21 nm. This compares to 1.18 nm obtained with the CG-MD and MZ-DPD models.

Dynamic Properties

Figure 6.11 shows the normalized molecular VACFs calculated from short trajectories:

$$c_{vv}(t) = \frac{\langle \mathbf{v}_{mol}(t) \cdot \mathbf{v}_{mol}(0) \rangle}{\langle \mathbf{v}_{mol}(0) \cdot \mathbf{v}_{mol}(0) \rangle}, \tag{6.25}$$

where $\mathbf{v}_{mol}$ is the velocity of the molecular center of mass. The behavior observed for the monomer and dimer is very similar in character. The CG-MD VACFs show a slower initial decay and minima and maxima of higher amplitude before decaying to zero in comparison to its FG-MD counterpart. The VACFs

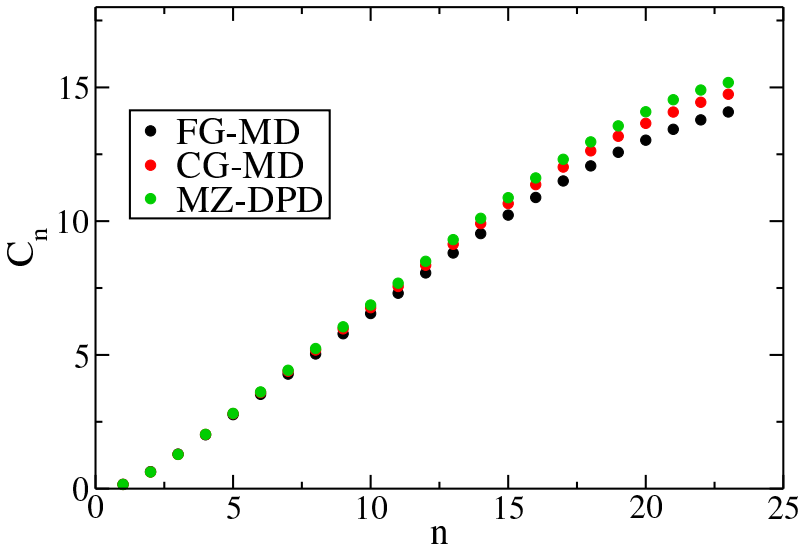


Figure 6.10: Characteristic ratio C_n of the 24mer. Small differences between CG-MD and MZ-DPD values at $n > 15$ are due to statistical inaccuracies.

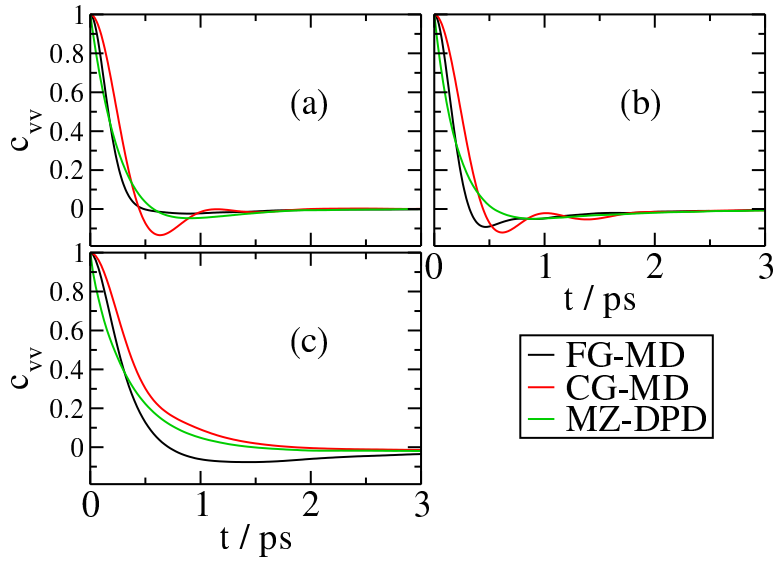


Figure 6.11: VACF of the molecular center of mass of the monomer (a), dimer (b), and 24mer (c) using the three different models.

Table 6.1: Self-diffusion coefficients D and rotational relaxation times τ_R of the pure liquids.

	D / 10^{-5} cm^2 s^{-1}			τ_R / ps	
method	Monomer	Dimer	24mer	Dimer	24mer
FG-MD	4.6	1.1	0.10	27	1790
CG-MD	7.2	2.0	0.45	12	620
MZ-DPD	5.1	1.3	0.23	17	1080

obtained with the MZ-DPD model are strongly dampened and decay on time scales similar to the those of the FG-MD VACFs. Despite improvement achieved in comparison with CG-MD, the FG-MD VACF is not reproduced with MZ-DPD for any of the three systems. This result is expected for systems with time scales that are not clearly separated and confirms findings from a previous study where a similar model of (mono-)neopentane has been used.[36]

It is interesting to note at this point that for very short times the MZ-DPD VACF always decays faster than the FG-MD VACF. For longer times, there is a crossover in this behavior that leads to the exact opposite trend. At very short times, the motion of the molecular CG system is ballistic in FG-MD while it is dissipative in MZ-DPD. Hence, molecular velocities decorrelate faster in MZ-DPD on this ballistic time scale. On longer time scales, particle collisions lead to further decay of the VACF. The MZ-DPD VACF now decays slower than the FG-MD VACF. While the MZ-DPD model assumes an adiabatic, Born-Oppenheimer-type, response of the Q-DoF in every particle collision, in reality (FG-MD) these Q-DoF dissipate less kinetic energy than assumed by the MZ-DPD model. Clearly, this happens because the time scale of the Q-DoF does not separate from the time scale of the P-DoF. Hence, collisions between mapped CG sites in FG-MD are "more elastic" than assumed in MZ-DPD, leading to a faster randomization of velocity vectors in FG-MD and a faster decay of the VACF. We however expect that, despite differences in the VACFs, diffusion coefficients are the same in the FG-MD and MZ-DPD simulations because the dissipative pair interaction applied in MZ-DPD is obtained in our approach by averaging over many realizations of pair collisions in the FG model.

In Table 6.1 we summarize the self-diffusion coefficients calculated from the mean squared displacement of the interaction sites using the Einstein method. The diffusion coefficients obtained from integrating the VACFs produce identical results. The data in Table 6.1 show that the diffusion coefficients obtained with the CG-MD models overestimate the ones obtained with the FG-MD model. With the MZ-DPD model, the monomer and dimer self-diffusion coefficients are almost exactly reproduced. This result agrees with earlier results[36] and shows that the MZ-DPD approach also produces a model with consistent long-time dynamics for a molecule containing two CG interaction sites. For the 24mer system, MZ-DPD model is not as successful. We find that D is about 2.3 times larger in the MZ-DPD simulation than in the FG-MD simulation. It is not obvious from the presented results, why the MZ-DPD model of the 24mer performs less well in reproducing the FG dynamics. It should be noted, however, that the method used herein to calculate the DPD frictions ignores multi-body contributions, e.g., originating from the immediate bonded environment of a bead in the polymer chain. Lemarchand et al. have recently shown that intramolecular contributions to the friction coefficients cannot be neglected.[35] Therefore, we expect that the inclusion of these contributions (e.g. within the framework of multiscale coarse-graining[6]) leads to an improved description of polymer dynamics with MZ-DPD. A further reason for the observed mismatch is related to energy barrier crossing, which, in the 24mer liquid, may necessitate the inclusion of explicit memory in the CG model.[49]

Figure 6.12 shows the time correlation function $C_{ee}(t)$ of the end-to-end vector orientation of the dimer and 24mer:

$$C_{ee}(t) = \langle \mathbf{e}_{ee}(0) \cdot \mathbf{e}_{ee}(t) \rangle, \tag{6.26}$$

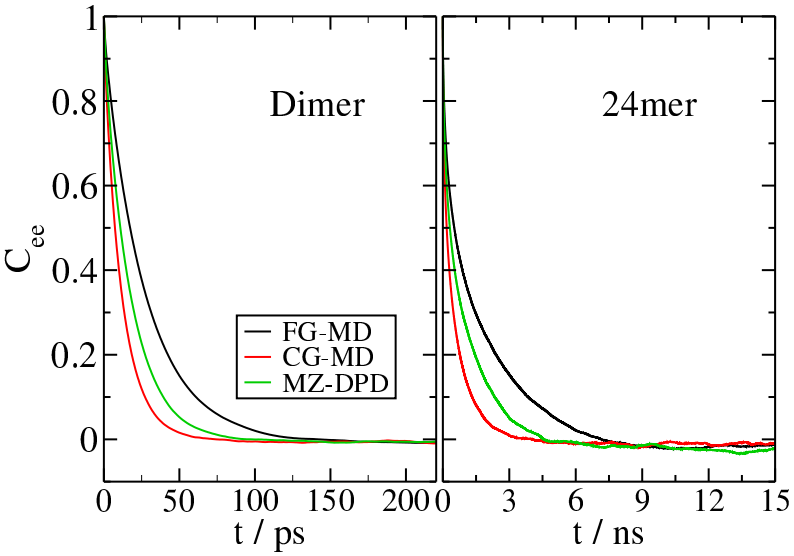


Figure 6.12: Auto correlation function of the unit vector parallel to the end-to-end distance vector.

where $\mathbf{e}_{ee}$ is a unit vector parallel to the connection vector between the first and last CG sites. The results obtained for the dimer and 24mer are qualitatively similar. Again, a speed-up of the CG-MD model with respect to the FG-MD model is observed which leads to a faster decay of $C_{ee}(t)$. The MZ-DPD model reduces the amount of this speed-up but does not succeed in reproducing the behavior of the FG-MD model. $C_{ee}(t)$ can be fitted using an exponential function to obtain the rotational relaxation time τ_R:

$$C_{ee}(t) = \exp\left(-\frac{t}{\tau_R}\right). \tag{6.27}$$

The values of τ_R are shown in Table 6.1. For both molecules a speed-up of the rotation in the CG-MD model is observed leading to a decrease in τ_R. For MZ-DPD model, τ_R is closer to the FG-MD result, but the relaxation times are still only 60% of the value observed in the FG-MD reference models.

In conclusion, the above results illustrate that the dynamics predicted with MZ-DPD models is consistent with fine-grained dynamics on diffusive time scales for the small molecular components despite incomplete time scale separation. This is not achieved for the 24mer, presumably due to missing multibody contributions in the DPD frictions obtained within the EF-CG framework.

6.6.3 Polymer-Solvent Mixtures

Coarse graining of polymer solutions, and multicomponent systems in general, may result in CG-MD models in which relative differences between dynamic properties of the individual components are inconsistent in comparison with FG models and experiments. Inconsistencies in the characteristic time scales for the motions of the different components are problematic when dynamic processes are studied in computer simulations. It is therefore interesting to compare the FG-MD, CG-MD and MZ-DPD models for multicomponent systems. To this end, we set up simple polymer solution models in which the 2,2-dimethyl propane 24mer represents the polymer and the 2,2-dimethyl propane dimer the solvent.

The VACFs of the molecular centers of mass of the molecule of both components in the mixtures P75S25 and P25S75 are shown in figure 6.13. In accordance with the VACFs of the pure liquids (Fig. 6.11), the MZ-DPD simulation of mixture P75S25 does not reproduce the FG dynamics on short time

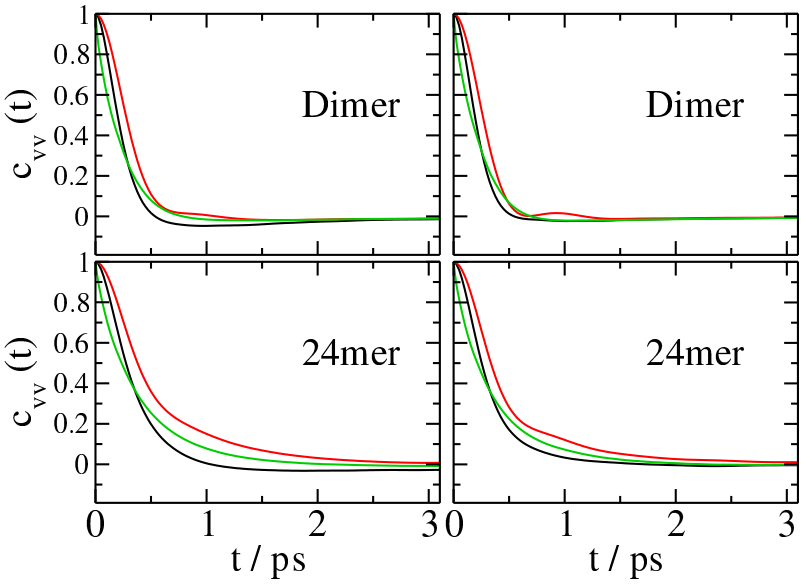


Figure 6.13: VACFs of the molecular centers of mass of the dimer and 24mer in mixture P75S25 (left) and mixture P25S75 (right). Black lines denote FG-MD, red lines CG-MD, and green lines MZ-DPD simulation results.

Table 6.2: Diffusion coefficients D of the components in polymer-solvent mixtures P75S25 and P25S75.

	D / 10^{-5} cm^2 s^{-1}		D / $D_{FG\text{-}MD}$	
method	P75S25	P25S75	P75S25	P25S75
Dimer				
FG-MD	2.4	3.1		
CG-MD	6.5	5.6	2.7	1.8
MZ-DPD	3.5	3.3	1.5	1.1
Polymer				
FG-MD	0.23	0.48		
CG-MD	0.89	0.90	3.9	1.9
MZ-DPD	0.41	0.42	1.8	0.9

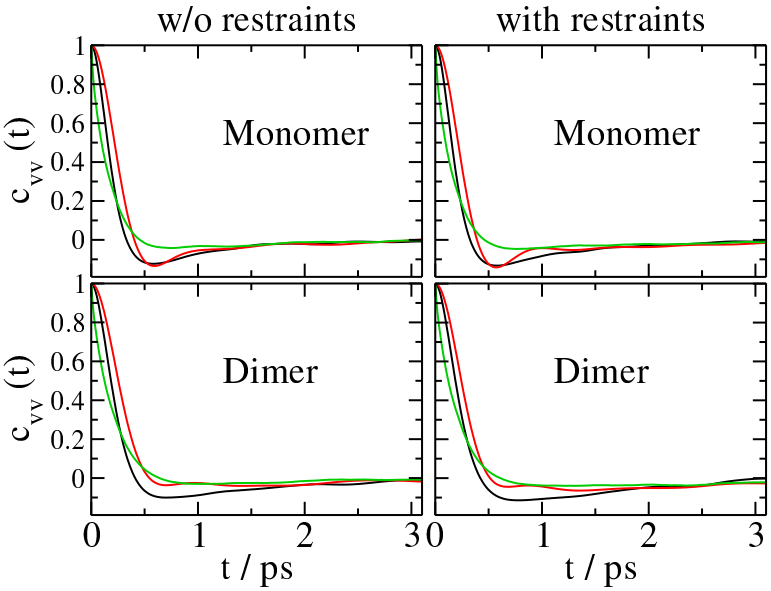


Figure 6.14: VACFs of the molecular centers of mass of the monomers and dimers in a poly-NEP matrix with (right panel) and without (left panel) restraints applied on the NEP repeat units. Black lines denote FG-MD, red lines CG-MD, and green lines MZ-DPD simulation results.

scales, although the agreement is better than the results obtained with CG-MD. For mixture P25S75 (with a larger solvent fraction), the MZ-DPD model performs better. It should however be noted that, for this system, the difference observed between the FG-MD and CG-MD VACFs is not as large as for the other mixture or for the pure component liquids (Fig. 6.11).

Table 6.2 summarizes the diffusion coefficients D of the solvent and polymer components in the two mixtures. We first consider mixture P75S25 (highest polymer content). When we compare the CG-MD and FG-MD models, we find that the diffusion ratio $D_{\text{CG-MD}}/D_{\text{FG-MD}}$ is different for the two mixture components: the solvent (dimer) diffusion is accelerated with a factor 2.7, while for the polymer diffusion the corresponding factor is 3.9. With the MZ-DPD model, the diffusion ratios $D_{\text{MZ-DPD}}/D_{\text{FG-MD}}$ are 1.5 for the solvent and 1.8 for the polymer. This not only means that the diffusion coefficients obtained with MZ-DPD are closer to FG-MD, but also that the ratio of the solvent and polymer diffusion coefficients obtained with MZ-DPD is in better agreement with the FG-MD model in comparison with CG-MD. The MZ-DPD model does not reproduce the FG-MD diffusion coefficients quantitatively. The largest discrepancy is observed between the polymer diffusion coefficients. This may again be related to missing multi-body contributions in the DPD frictions between polymer interaction sites, as discussed above.

We next consider mixture P25S75 (higher dimer content) in Table 6.2. The diffusion ratios $D_{\text{CG-MD}}/D_{\text{FG-MD}}$ are smaller than for mixture P75S25 and are approximately the same (1.8 and 1.9) for both components. This is not suprising if one considers that diffusion of dilute polymers is determined by the viscosity of the solvent (Stokes-Einstein). When we increase the solvent diffusion coefficient with a factor 1.8 (CG-MD), we decrease the solvent viscosity and, therefore, increase polymer diffusion with approximately the same factor. Significantly, the MZ-DPD model closely reproduces the solvent diffusion coefficient of the FG-MD model. For the polymer, almost equally good agreement is achieved.

6.6.4 Penetrant Diffusion in a Polymer Matrix

As a final application of the MZ-DPD models derived herein, we examine the dynamics of small penetrant molecules (monomers and dimers of 2,2-dimethyl propane) in a polymer matrix composed of long poly-

Table 6.3: Diffusion coefficients of monomers and dimers in the poly-NEP matrix

	D / 10^{-5} cm^2 s^{-1}		D / D_{FG-MD}		$D_{monomer}$ / D_{dimer}
method	Monomer	Dimer	Monomer	Dimer	
without restraints					
FG-MD	0.89	0.41			2.2
CG-MD	4.6	3.2	5.2	7.8	1.4
MZ-DPD	3.9	2.2	4.3	5.4	1.8
with restraints					
FG-MD	0.46	0.06			7.7
CG-MD	1.2	0.83	2.6	14	1.4
MZ-DPD	0.77	0.35	1.7	5.8	2.2

NEP chains. The penetrant diffusion mechanism in polymer melts and glasses is governed by activated barrier crossing, or "hopping",[50, 51] rather than by molecular collisions that govern diffusion in liquids. The relevant energy barriers for penetrant hopping in polymers are determined by the conservative part of the interactions and tend to be underestimated with CG-MD models.[15, 52, 53] Therefore, quantitative prediction of penetrant diffusion in polymers requires CG models that capture the relevant features of, both, conservative and dissipative interactions. We will here examine the role of dissipative interactions by comparing CG-MD and MZ-DPD models in which the conservative forces are identical.

The VACFs of the molecular center of mass of both penetrant molecules are shown in figure 6.14. It is interesting to compare these data with the monomer/dimer VACFs in figures 6.11 and 6.13. Clearly, the agreement achieved between the MZ-DPD and FG-MD models in figure 6.14 is not as clear as for the other systems (figures 6.11 and 6.13) where dissipative molecular collisions, rather than activated barrier crossings, dominate the dynamics. This is confirmed by the comparison of FG-MD, CG-MD, and MZ-DPD penetrant diffusion coefficients D shown in Table 6.3. Interestingly, the improvement achieved with MZ-DPD over CG-MD is better for the polymer matrix simulated with restraints. This observation probably reflects an aspect of coupling between penetrant and chain motions: because the MZ-DPD model overestimates the dynamics of the polymer (Fig. 6.12), penetrant diffusion is overestimated in the unrestrained polymer matrix with MZ-DPD, too. When restraints are applied, we observe that the MZ-DPD provides better agreement. In the system with restraints, the diffusion coefficient ratio $D/D_{\text{FG-MD}}$ decreases from 2.6 to 1.7 (−35%) for the monomer when DPD frictions are applied; for the dimer, the diffusion ratio decreases from 14 to 5.8 (−59%). This compares with −17% (monomer) and −31% in the system without restraints.

The FG-MD penetrant diffusion coefficients are not recovered with MZ-DPD. This may be caused by inaccuracies in the CG description of energy barriers as mentioned above, but may furthermore be caused by memory effects as recently shown by Kappler et al.[49] These authors observed a memory-induced slowdown of barrier crossing under conditions that the memory time exceeds the diffusion time. For the present application of penetrant diffusion in polymers, this condition requires memory times of the order of nanoseconds corresponding to typical penetrant hopping times. If we assume that memory times for penetrant diffusion are long, and furthermore assume that they are longer for dimer diffusion than for monomer diffusion, this may explain why the ratio $D/D_{\text{FG-MD}}$ in Table 6.3 is larger for the dimer than for the monomer. It may also explain why the ratio $D_{\text{monomer}}/D_{\text{dimer}}$ in MZ-DPD does not match FG-MD.

6.7 Conclusions

We have used the effective-force coarse-graining (EF-CG) method in conjunction with the Q-approximation to derive Mori-Zwanzig-Dissipative-Particle-Dynamics (MZ-DPD) multiple-site CG models

from unconstrained fine-grained molecular dynamics simulation trajectories of 2,2-dimethyl-propane-based monomers, dimers, and 24mers. In our Markovian DPD implementation, we used a pairwise approximation for the conservative forces combined with the assumption that multi-body contributions in the DPD pair frictions are negligible. While time scale separation does not strictly apply to the molecular CG models examined herein, we show that the diffusive dynamics in pure liquids and in a polymer-solvent mixture with excess solvent can be accurately represented.

We find that the representability of polymer diffusion in concentrated polymer solutions and melts is less accurate compared with diffusion of single-site and two-site molecular liquids. This is presumably caused by multi-body correlations between bonded interaction sites in the multiple-bead CG representation of the polymer which are neglected in our model. Furthermore, the presence of energy barriers for dynamics may lead to stronger memory effects in these systems. In all systems examined, the time correlation function of the pair force fluctuations (FACF) exhibits a slowly decaying tail due to incomplete separation of time scales. The initial fast decay of the FACF is related to the relaxation of atomic vibrations, while the slowly decaying tail is related to molecular reorientations. The DPD pair frictions associated with the atomic vibrations were obtained by removing the long time tail from the FACF.

The model presented herein is successful for liquids but does not provide an accurate description of long-time dynamics governed by barrier crossing. We illustrated this with an application to diffusion of small molecular penetrants in a polymer network and found that the MZ-DPD approach overestimates penetrant diffusion coefficients. For this application, and, more generally, for systems where dynamics is governed by infrequent barrier crossing events rather than by dissipative particle collisions, inclusion of explicit memory may lead to improved descriptions. Recently, Kappler et al. for example found that memory-induced slowdown of barrier crossing occurs when the memory time exceeds the diffusion time.[49] An efficient implementation of explicit memory was recently reported by Li et al.[29] who used auxiliary variables to replace non-Markovian dynamics with Markovian dynamics in a higher dimensional space.[54]

The present work shows that a bottom-up parameterized DPD thermostat permits to represent long-time dynamic properties of multicomponent CG fluids models. Therefore, the MZ-DPD approach extends the representability of molecular CG force fields beyond structural and thermodynamic properties commonly employed to parameterize the conservative interactions. The approach is not limited to bottom-up EF-CG models and can be applied using conservative forces from other bottom-up or top-down CG models as well.

6.8 References

1. Peter, C. & Kremer, K. Multiscale simulation of soft matter systems. *Faraday Discussions* **144.** 00053, 9 (2010).
2. Brini, E. *et al.* Systematic coarse-graining methods for soft matter simulations – a review. *Soft Matter* **9.** 00021, 2108 (2013).
3. Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* **139,** 090901 (2013).
4. Lyubartsev, A. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E* **52.** 00390, 3730–3737 (1995).
5. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry* **24.** 00428, 1624–1636 (2003).
6. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics* **128,** 244114 (2008).
7. Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics* **129,** 144108 (2008).

8. Wang, Y., Noid, W. G., Liu, P. & Voth, G. A. Effective force coarse-graining. *Physical Chemistry Chemical Physics* **11,** 2002 (2009).

9. Brini, E., Marcon, V. & Van der Vegt, N. F. A. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics* **13.** 00022, 10468 (2011).

10. Tschöp, W., Kremer, K., Batoulis, J., Bürger, T. & Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica* **49.** 00310, 61–74 (1998).

11. Akkermans, R. L. C. & Briels, W. J. Coarse-grained dynamics of one chain in a polymer melt. *The Journal of Chemical Physics* **113,** 6409 (2000).

12. Izvekov, S. & Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *The Journal of Chemical Physics* **125,** 151101 (2006).

13. Hijón, C., Español, P., Vanden-Eijnden, E. & Delgado-Buscalioni, R. Mori–Zwanzig formalism as a practical computational tool. *Faraday Discussions* **144,** 301 (2010).

14. Fritz, D., Koschke, K., Harmandaris, V. A., Van der Vegt, N. F. A. & Kremer, K. Multiscale modeling of soft matter: scaling of dynamics. *Phys. Chem. Chem. Phys.* **13,** 10412 (2011).

15. Fritz, D., Herbers, C. R., Kremer, K. & Van der Vegt, N. F. A. Hierarchical modeling of polymer permeation. *Soft Matter* **5,** 4556 (2009).

16. Harmandaris, V. A. & Kremer, K. Dynamics of Polystyrene Melts through Hierarchical Multiscale Simulations. *Macromolecules* **42,** 791–802 (2009).

17. Zwanzig, R. Memory Effects in Irreversible Thermodynamics. *Physical Review* **124,** 983–992 (1961).

18. Mori, H. Transport, Collective Motion, and Brownian Motion. *Progress of Theoretical Physics* **33,** 423–455 (1965).

19. Kawasaki, K. Simple derivations of generalized linear and nonlinear Langevin equations. *Journal of Physics A: Mathematical, Nuclear and General* **6,** 1289–1295 (1973).

20. Kinjo, T. & Hyodo, S. Equation of motion for coarse-grained simulation based on microscopic description. *Physical Review E* **75,** 051109 (2007).

21. Hoogerbrugge, P. J. & Koelman, J. M. V. A. Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics. *Europhysics Letters (EPL)* **19.** 01901, 155–160 (1992).

22. Español, P. & Warren, P. Statistical Mechanics of Dissipative Particle Dynamics. *Europhysics Letters (EPL)* **30.** 00000, 191–196 (1995).

23. Groot, R. D. & Warren, P. B. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation. *The Journal of Chemical Physics* **107.** 01852, 4423 (1997).

24. Español, P. & Warren, P. B. Perspective: Dissipative particle dynamics. *The Journal of Chemical Physics* **146,** 150901 (2017).

25. Eriksson, A., Jacobi, M. N., Nyström, J. & Tunstrøm, K. A method for estimating the interactions in dissipative particle dynamics from particle trajectories. *Journal of Physics: Condensed Matter* **21.** 00007, 095401 (2009).

26. Li, Z., Bian, X., Caswell, B. & Karniadakis, G. E. Construction of dissipative particle dynamics models for complex fluids via the Mori–Zwanzig formulation. *Soft Matter* **10,** 8659–8672 (2014).

27. Lei, H., Caswell, B. & Karniadakis, G. E. Direct construction of mesoscopic models from microscopic simulations. *Physical Review E* **81.** 00039, 026704 (2010).

28. Li, Z., Bian, X., Li, X. & Karniadakis, G. E. Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism. *The Journal of Chemical Physics* **143,** 243128 (2015).

29. Li, Z., Lee, H. S., Darve, E. & Karniadakis, G. E. Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts. *The Journal of Chemical Physics* **146,** 014104 (2017).

30. Eriksson, A., Jacobi, M. N., Nyström, J. & Tunstrøm, K. Effective thermostat induced by coarse graining of simple point charge water. *The Journal of Chemical Physics* **129.** 00017, 024106 (2008).

31. Eriksson, A., Jacobi, M. N., Nyström, J. & Tunstrøm, K. Bottom-up derivation of an effective thermostat for united atoms simulations of water. *The Journal of Chemical Physics* **130.** 00008, 164509 (2009).

32. Izvekov, S. & Rice, B. M. Multi-scale coarse-graining of non-conservative interactions in molecular liquids. *The Journal of Chemical Physics* **140.** 00000, 104104 (2014).

33. Trément, S., Schnell, B., Petitjean, L., Couty, M. & Rousseau, B. Conservative and dissipative force field for simulation of coarse-grained alkane molecules: A bottom-up approach. *The Journal of Chemical Physics* **140,** 134113 (2014).

34. Dequidt, A. & Solano Canchaya, J. G. Bayesian parametrization of coarse-grain dissipative dynamics models. *The Journal of Chemical Physics* **143,** 084122 (2015).

35. Lemarchand, C. A., Couty, M. & Rousseau, B. Coarse-grained simulations of *cis* - and *trans* - polybutadiene: A bottom-up approach. *The Journal of Chemical Physics* **146,** 074904 (2017).

36. Deichmann, G., Marcon, V. & Van der Vegt, N. F. A. Bottom-up derivation of conservative and dissipative interactions for coarse-grained molecular liquids with the conditional reversible work method. *The Journal of Chemical Physics* **141,** 224109 (2014).

37. Izvekov, S. Microscopic derivation of particle-based coarse-grained dynamics. *The Journal of Chemical Physics* **138,** 134106 (2013).

38. Yoshimoto, Y. *et al.* Bottom-up construction of interaction models of non-Markovian dissipative particle dynamics. *Physical Review E* **88,** 043305 (2013).

39. Jung, G., Hanke, M. & Schmid, F. Iterative Reconstruction of Memory Kernels. *Journal of Chemical Theory and Computation* **13,** 2481–2488 (2017).

40. Izvekov, S. & Rice, B. M. On the importance of shear dissipative forces in coarse-grained dynamics of molecular liquids. *Physical Chemistry Chemical Physics* **17,** 10795–10804 (2015).

41. Gao, L. & Fang, W. Semi-bottom-up coarse graining of water based on microscopic simulations. *The Journal of Chemical Physics* **135,** 184101 (2011).

42. Izvekov, S. Mori-Zwanzig theory for dissipative forces in coarse-grained dynamics in the Markov limit. *Physical Review E* **95,** 013303 (2017).

43. Izvekov, S. Microscopic derivation of particle-based coarse-grained dynamics: Exact expression for memory function. *The Journal of Chemical Physics* **146,** 124109 (2017).

44. Martin, M. G. & Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *The Journal of Physical Chemistry B* **102,** 2569–2577 (1998).

45. Hess, B., Kutzner, C., Van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **4,** 435–447 (2008).

46. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **117.** 05802, 1–19 (1995).

47. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52.** 04126, 255–268 (1984).

48. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

49. Kappler, J., Daldrop, J. O., Brünig, F. N., Boehle, M. D. & Netz, R. R. Memory-induced acceleration and slowdown of barrier crossing. *The Journal of Chemical Physics* **148,** 014903 (2018).

50. Gusev, A. A. & Suter, U. W. Dynamics of small molecules in dense polymers subject to thermal motion. *The Journal of Chemical Physics* **99,** 2228–2234 (1993).

51. Müller-Plathe, F. Permeation of polymers — a computational approach. *Acta Polymerica* **45,** 259–293 (1994).

52. Depa, P. K. & Maranas, J. K. Speed up of dynamic observables in coarse-grained molecular-dynamics simulations of unentangled polymers. *The Journal of Chemical Physics* **123,** 094901 (2005).

53. Depa, P. K. & Maranas, J. K. Dynamic evolution in coarse-grained molecular dynamics simulations of polyethylene melts. *The Journal of Chemical Physics* **126,** 054903 (2007).

54. Ceriotti, M., Bussi, G. & Parrinello, M. Colored-Noise Thermostats à la Carte. *Journal of Chemical Theory and Computation* **6,** 1170–1180 (2010).

7 Conclusions and Outlook

In the preceding chapters, four studies have been presented highlighting the qualities of the pair-interaction based systematic coarse-graining methods EF-CG and CRW.

In chapter 3, it has been shown that, in addition to the quantities investigated in previous studies, CG models parameterized with these methods represent the vapor-liquid phase equilibria of hexane and perfluorohexane well in comparison with the structure-based IMC models. The results in this chapter illustrate the transferability of the pair-interaction based CG models with respect to the temperature, the environment change associated with moving a molecule from the bulk to the interface, and the composition of a binary mixture of the two components. The differences observed between the EF-CG and CRW models are significant, which is somewhat surprising given the similarity of the EF-CG and CRW potentials, and it is not completely clear yet what the origin of these differences is.

Possible causes are, first, the surroundings of the pair of atom groups used for the CG potential parameterization and, second, the slight differences in the parameterization algorithms. Further insight into the differences between EF-CG and CRW methods might be obtained by comparing potentials which have been parameterized by both methods at the same state point and in the same sampling environment. CRW parameterization can be performed in the liquid[1] and it is also straightforward to perform EF-CG parameterization in the same setup as the standard CRW parameterization, using two molecules in vacuo. A systematic comparison of different surroundings and mappings might illustrate whether the differences in the calculation of the "excluded interaction" (discussed in section 2.4.2) are really significant and how such differences depend on the mapping chosen for the CG model.

In the two following chapters (4 and 5), a modified CRW algorithm is presented to separate the CG interaction potential into two effective components: a Van-der-Waals (VdW) and an electrostatic (ES) contribution. This decomposition of the interaction free energy allows for a better identification of possible simplifications in the model. As a consequence, the ES terms of the CG models have been approximated using point charges with the magnitude of the net charge sum of the mapped FG system. Further, the approximation of the VdW contribution through a functional form significantly reduces the number of parameters required for the description of the CG force field and allows for fine-tuning to enhance the representability of the CG model at the state point of parameterization. The qualities of CRW models for systems with polar and ionic liquids are similar to those observed in studies of apolar liquids: the transferability of these models is good but this quality comes at the expense of representability especially w.r.t. the liquid structure. While the thermodynamics of the systems is reproduced to the degree observed in earlier studies,[1–4] the electric properties are not well reproduced by the CRW CG models. The reasons for this deficiency require more investigation into the mapping schemes and the correct distribution of the atomistic charges on the CG interactions sites. In a first step, mappings are required that preserve the dipole distribution of the molecule also in the CG representation. Then, investigations can be undertaken concerning the relation between the CG interaction potentials and the electric properties of the liquid resulting from a given model.

In the last chapter, a method for dynamically consistent CG modeling of CG systems with molecular mappings has been presented. Building on previous applications of the Mori-Zwanzig projection operator formalism in the literature, a bottom-up MZ-DPD model has been constructed based on the EFCG coarse-graining method. The focus here is on the dependence of the dynamic consistency on the mapping and the question whether MZ-DPD models can be calculated for systems with molecular mappings using the existing techniques published in the literature. To this end, three molecules containing a different number of CG sites of the same size (5 CH_x units) were modeled. The results for the smaller molecules confirm earlier findings for MZ-DPD models at comparatively high density and high resolution:[5–7] the dynamical behavior on diffusive time scales is well reproduced, whereas the relaxation of short time dynamics is inconsistently represented in the MZ-DPD model. In comparison to the FG reference

dynamics, the velocity auto-correlation function (VACF) is dampened stronger and does not show a pronounced minimum. The reasons of this failure are probably related to the approximations made in the course of the MZ projection operator formalism. Herein it is assumed that the time scales between the relaxation of velocities and forces on a CG site are clearly separated which is not the case in a typical soft matter system at realistic density and low to medium degree of coarse-graining. Therefore, it is unlikely that the simple DPD EOM is suitable to simulate the short time dynamics in accordance with a FG reference for the model systems utilized in this work. A way forward may be the utilization of computationally efficient algorithms including memory effects, i.e. the explicit time-dependence of the pair friction coefficients.[8] Whether such a simulation with explicit memory is still computationally efficient at high density and with molecular CG mappings is not clear yet and has to be evaluated in further research.

In the homogeneous liquids studied here, the quality of reproduction of the self-diffusion coefficient decreases with increasing chain length. This can be regarded as an indication that the internal friction, which is neglected in the utilized MZ-DPD models is actually an important component for a dynamically consistent representation of a system with a molecular mapping as pointed out in another recent study.[9]

In addition to the homogeneous liquid systems, dynamic properties of a system of 24mers and mixtures of low molecular weight liquids and this polymer are investigated. Here it has been shown, that also the diffusive dynamics of the FG reference cannot be reproduced in CG-DPD simulations for the 24mer system. The CG potential energy surface of this system is softened in such a manner that the mechanism of penetrant diffusion in the polymer is significantly altered. New routes open up for the penetrant to move within the polymer matrix that were not accounted for in the parameterization of the MZ-DPD pair friction coefficients. This effect cannot be compensated for by the addition of bottom-up friction terms alone. For a realistic CG model of penetrant diffusion, also the energy barriers of the diffusion process have to be realistically modeled.

7.1 References

1. Brini, E. & Van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *The Journal of Chemical Physics* **137.** 00009, 154113 (2012).
2. Fritz, D., Harmandaris, V. A., Kremer, K. & Van der Vegt, N. F. A. Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities. *Macromolecules* **42.** 00065, 7579–7588 (2009).
3. Brini, E., Herbers, C. R., Deichmann, G. & Van der Vegt, N. F. A. Thermodynamic transferability of coarse-grained potentials for polymer–additive systems. *Physical Chemistry Chemical Physics* **14.** 00006, 11896 (2012).
4. Dallavalle, M. & Van der Vegt, N. F. A. Evaluation of mapping schemes for systematic coarse graining of higher alkanes. *Phys. Chem. Chem. Phys.* **19,** 23034–23042 (2017).
5. Trément, S., Schnell, B., Petitjean, L., Couty, M. & Rousseau, B. Conservative and dissipative force field for simulation of coarse-grained alkane molecules: A bottom-up approach. *The Journal of Chemical Physics* **140,** 134113 (2014).
6. Izvekov, S. & Rice, B. M. Multi-scale coarse-graining of non-conservative interactions in molecular liquids. *The Journal of Chemical Physics* **140.** 00000, 104104 (2014).
7. Deichmann, G., Marcon, V. & Van der Vegt, N. F. A. Bottom-up derivation of conservative and dissipative interactions for coarse-grained molecular liquids with the conditional reversible work method. *The Journal of Chemical Physics* **141,** 224109 (2014).
8. Li, Z., Bian, X., Li, X. & Karniadakis, G. E. Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism. *The Journal of Chemical Physics* **143,** 243128 (2015).

9. Daldrop, J. O., Kappler, J., Brünig, F. N. & Netz, R. R. Butane dihedral angle dynamics in water is dominated by internal friction. *Proceedings of the National Academy of Sciences* **115,** 5169–5174 (2018).